Title: Gut microbiome meta-analysis reveals dysbiosis is independent of body mass index in predicting risk of obesity-associated CRC

Authors: K. Leigh Greathouse[1,4], James Robert White[2], R. Noah Padgett[3], Brittany G Perrotta[4], Gregory D Jenkins[5], Nicholas Chia[5,6], Jun Chen[6].

Author Affiliations: 1 Nutrition Sciences Division, Robbins College of Health and Human Science, Baylor University, Waco, TX; 2 Resphera Biosciences, Baltimore, MD, USA; 3 Department of Educational Psychology, Baylor University; 4 Department of Biology, Baylor University; 5 Department of Surgery and Department of Physiology and Biomedical Engineering, Mayo Clinic; 6 Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic.

Abstract:

Obesity is a risk factor for colorectal cancer (CRC), accounting for more than 14% of CRC incidence. Microbial dysbiosis and chronic inflammation are common characteristics in both obesity and CRC. Human and murine studies, together, demonstrate the significant impact of the microbiome on governing energy metabolism and CRC development; yet, little is understood about the contribution of the microbiome to development of obesity-associated CRC as compared to non-obese individuals. In this study, we conducted a meta-analysis using five publicly available stool and tissue-based 16S rRNA and whole genome sequencing (WGS) data sets of CRC microbiome studies. High-resolution analysis was employed for 16S rRNA data using Resphera Insight, which allowed us to achieve species-level information to compare with WGS. Characterization of the confounders between

studies, 16S rRNA variable region, and sequencing method, did not reveal any significant effect on alpha diversity in CRC prediction. Both 16S rRNA and WGS were equally variable in their ability to predict CRC. Results from community structure and composition analysis confirmed lower diversity in obese individuals without CRC; however, no universal differences were found in diversity between obese and non-obese individuals with CRC. When examining taxonomic differences, the probability of being classified as CRC did not change significantly in obese individuals for all taxa tested. However, random forest classification was able to distinguish CRC and non-CRC stool when body mass index was added to the model. Overall, microbial dysbiosis was not a significant factor in explaining the higher risk of colon cancer among individuals with obesity.

Introduction. The percentage of individuals who are overweight or obese in the U.S. has reached epidemic proportions, with the prevalence of individuals who are overweight (32.7%) or obese (34.3%), as defined by body mass index (BMI), in the United States representing about two thirds of adult Americans. The health risks associated with overweight and obesity include diabetes, cardiovascular disease, and cancer. The National Cancer Institute estimates that 3.2% of all new cancers are due to obesity and that 14% of deaths from cancer in men and 20% in women are attributed to obesity (1, 2). Colorectal cancer (CRC) accounts for approximately 142,000 new cancer cases and 50,000 cancer deaths annually, making it the second most lethal cancer in the U.S. (SEER). Several epidemiological studies demonstrate that adult obesity increases the risk of colon cancer 1.2 to 2-fold, with obesity accounting for 14-35% of total colon cancer

incidence (1, 3-5). Alarmingly, incidence and mortality from CRC is on the rise among those under the age of 55 (SEER), possibly due to the significant increase in obesity in women (6). For these reasons, it is imperative to identify new methods to reduce the burden of obesity on the risk and mortality from colon cancer. Three areas of inquiry are important for understanding the etiology of CRC: obesity, inflammation, the microbiome.

Several studies indicate that specific microbial taxa are playing a role in the etiology of colon cancer. However, whether the microbiome is also contributing to development of obesity-associated colon cancer in *humans* is completely unknown. One method that has shown promise for identifying early stage colon cancer is through analyzing the microbiome of the gastrointestinal tract (GI). The structure and function of the bacterial community that makes up the human colon, in part, determines the function and health of the colonic epithelium, as well as, the immune system responses. Several studies have found colon cancer-associated microbiota in pre-cancerous colon tissue (adenomas). Further, the microbiome has been used to distinguish pre-cancerous adenomas from CRC, though with variable rates of accuracy (7-9). Several bacteria have been identified as promoters in colon cancer development, including enterotoxigenic *Bacteroides fragilis* and *Fusobacterium nucleatum* (10-14). Both have also been isolated from patients with familial adenomatous polyposis (FAP) or inflammatory bowel disease, which are risk factors for colon cancer (11, 15). Colorectal adenocarcinomas associated with high abundance of fecal *F. nucleatum,* specifically, were found to have the highest number of somatic mutations, suggesting that these mutations create a pathogen-

friendly environment (16, 17). In animal models of colon cancer, inoculation of germ-free animals with stool from tumor-bearing animals were found to have more tumors than mice inoculated with stool from tumor-free animals (18). More recently, colonic biofilms from individuals with familial adenomatous polyposis were found to be dominated by *E. coli* and *B.* fragilis biofilms and enriched with genotoxic colibactin and *B. fragilis* toxin (ETBF) genes (11).

Multiple lines of evidence demonstrate that both diet and obesity can significantly alter the microbiome (19-25). One of the first seminal studies illustrating the impact of the microbiome on obesity, transferred the fecal microbiota from monozygotic twins who were obese or lean to germ-free mice. From this study, they were able to recapitulate the obesity phenotype in humanized mice (26). When examining microbiota and subsequent changes in metabolism after fecal transfer from obese mice to germ-free mice, it was found that this obesogenic microbial community had an increased production of SCFAs, which was later shown to abrogate lipid storage (23, 26). Multiple follow-up studies in obese and lean individuals have linked the specific shift in the microbiota to the ratio of *Bacteroides:Firmicutes* (25, 27, 28). However, a recent meta-analysis of these studies indicate that this ratio is not sufficient to differentiate obese from lean individuals in separate studies. Thus, more research is necessary to identify the microbiome- host relationship in individuals with obesity (29, 30).

Chronic inflammation is a hallmark of both obesity and CRC etiology. Obesity is characterized by pro-inflammatory adipose tissue macrophages that secrete high levels of IL-17, a cytokine which is also induced by ETBF in murine models of colon

cancer (10, 31, 32). Given the reciprocal relationship between the microbiome and the immune system, it is logical to hypothesize that obesity-associated microbial dysbiosis, combined with a state of chronic inflammation, contributes to the increased risk of colon cancer among obese individuals. In support of this hypothesis, animal models of colon cancer (Apc^{1638N}), have demonstrated that a high fat diet or genetically (ob/ob) induced obesity can significantly alter the microbiome leading to a loss of *Parabacteroides distasonis* and an increase in pro- inflammatory factors (22). In a separate model of colon cancer ($K\text{-}ras^{G12Dint}$), fecal transfer from high-fat fed mice with intestinal tumors to genetically susceptible mice on a standard diet replicated the disease phenotype (33). Thus, it appears that a high fat diet may be sufficient to change the microbiome into a tumor-promoting community independent of obesity and glucose response. Intriguingly, *Akkermansia muciniphila,* which is reduced in obese individuals and is associated with epithelial barrier function, is paradoxically higher in CRC (34, 35). This data, together with evidence that *A. muciniphila* can modulate glucose metabolism and inflammation in the colon, suggests it may play a role in obesity-associated CRC (35, 36). As these data demonstrate, there are a variety of dysbiotic states that exist in obese individuals, which could further enhance the inflammatory state of the GI tract leading to an increased risk of CRC. No human studies to date have addressed the obesity-associated differences in the microbiome and its relationship to CRC however.

In this study, we utilized multiple publicly available data sets in which either stool or tissue microbiome sequencing was conducted, and from which body mass index (BMI) was also available. Using the bioinformatics tools QIIME (16S rRNA) and

Pathoscope (WGS), we processed the 16S rRNA and WGS reads, and derived a taxonomic profile from each of the samples. Furthermore, we inferred taxonomic function to assess potential metabolic differences in obese individuals with CRC. We used these taxa and the metabolic pathway information to determine if a taxonomic signature or if specific taxa were associated with both obesity and CRC. From this analysis, we observed that the dysbiosis associated with obesity was independent from the dysbiosis associated with CRC.

METHODS

Sample Population

For this study we identify studies relevant to assess the relationship between obesity and CRC using the microbiome as the independent variable using the following search terms in PubMed (((((("humans"[MeSH Terms] AND ("2006"[PDAT] : "2016"[PDAT])) NOT Review[Publication Type]) AND (obesity[Text Word] OR bmi[Text Word] OR body mass index[Text Word] OR BMI[Text Word] OR obesity[Text Word])) AND (bacterial[All Fields] AND ("microbiota"[MeSH Terms] OR "microbiota"[All Fields] OR "microbiome"[All Fields]))) AND (("colonic neoplasms"[MeSH Terms] OR ("colonic"[All Fields] AND "neoplasms"[All Fields]) OR "colonic neoplasms"[All Fields] OR ("colon"[All Fields] AND "cancer"[All Fields]) OR "colon cancer"[All Fields]) OR ("colorectal neoplasms"[MeSH Terms] OR ("colorectal"[All Fields] AND "neoplasms"[All Fields]) OR "colorectal neoplasms"[All Fields] OR ("colorectal"[All Fields] AND "cancer"[All Fields]) OR "colorectal cancer"[All Fields]) OR CRC[All Fields]) . From our PubMed search we identified 5 (out of 124) studies that met all of our criteria: primary research in a human population, colon or colorectal cancer AND

normal stool or tissue collected, raw sequences available from either 16S rRNA or WGS sequencing, body mass index available as a variable in the metadata including age and sex. Together, 5 studies were identified that assessed both BMI and the microbiome in stool or tissue from individuals with adenomas, carcinomas or individuals without disease (Table 1). Three of these studies conducted 16S rRNA sequencing on stool or tissue, and 3 conducted WGS on stool or tissue, with one utilizing RNA sequencing. One study conducted both 16S rRNA and WGS on tissue and stool.

Processing of Microbial Reads and Calculation of Diversity

All sequence data were downloaded from the NCBI Sequence Read Archive. In order to eliminate differences between studies, we processed the reads using the same methods, either QIIME plus the algorithm Resphera Insight for 16S rRNA sequencing or Pathoscope (v1.0) for processing WGS or RNA-seq reads. For the studies sequencing the 16S rRNA gene, the V4 region was used for all stool samples, as well as, tissue, with the exception of the subsample of tissue from another study used as part of the Zeller et al. 2014 data set. Details regarding sequencing methods and variable regions amplified for each data set are listed in Table 1.

Raw paired-end reads reflecting 16S rRNA fragments were merged into consensus sequences using FLASH (min overlap: 20 bp overlap; 5% max mismatch density), and trimmed for quality (target error rate < 1%) using Trimmomatic and QIIME. PhiX control sequences were identified using BLASTN and filtered. Resulting sequences were evaluated for chimeras with UCLUST (*de novo* mode) and screened for human DNA using Bowtie2 against NCBI *Homo sapiens* Annotation Release 106. Reads assigned to

chloroplast or mitochondrial contaminants by the RDP classifier with a minimum confidence of 50% were also removed. High-quality 16S rRNA sequences were assigned to a high-resolution taxonomic lineage using Resphera Insight (37-39) Raw paired-end shotgun metagenomics sequence datasets were also trimmed for quality using Trimmomatic (min final length 75bp) and screened for human genomic DNA using Bowtie2 (--sensitive setting against GRCh38 reference with alternate chromosomes). High-quality passing sequences were submitted to Pathoscope v1.0 for species level characterization (40, 41).

Prediction of Metagenomic Pathways

We utilized two methods in order to derive abundance of metabolic pathways from the 16S rRNA or WGS sequences. For the 16S rRNA reads, after obtaining the OTU tables, we utilized the PICRUSt algorithm. This method obtains the representative genomes according the nearest neighbor match, and then normalizes the genome abundance using the 16S rRNA copy number for that genome. Once the metagenomics content is binned, it is expressed in terms of KEGG representative ortholog (KO) counts. For the WGS reads, we utilized the HUMAnN algorithm. This method takes as input short DNA or RNA reads and uses BLAST to identify orthologous gene families, which are used to identify metabolic pathways. Once identified, the pathways are then normalized by presence/absence of the taxa, and additionally by relative abundance of the taxa present in the sample. These data were used for downstream statistical analysis to compare obese and normal stool samples from individuals with or without CRC.

Statistical Analyses

Prior to analysis we rarefied the data to the sample with the lowest number of reads. In order to test the association between BMI and the microbiome, we grouped our statistical analyses into four subgroups: A) normal stool samples (healthy controls), B) CRC stool or CRC tissue, and C) pooled samples (healthy controls and CRC), all of which were adjusted for age and sex. Group C was further adjusted for disease status.

For alpha diversity measurements, we used both the observed number of OTUs and the Shannon Index. To determine associations with BMI, we treated it as a continuous variable (as a covariate) in the main analysis. For additional analyses, we also dichotomized the subjects into non-obese (BMI < 30) and obese (BMI >= 30) according the WHO guidelines.

For beta diversity measurements, we utilized four distance measurements unweighted UniFrac, weighted UniFrac, generalized UniFrac and Bray-Curtis for 16S datasets. For WGS/RNA-seq data, where we do not have the phylogenetic tree, we instead used two non-tree-based distance measurements Jensen-Shannon and Bray-Curtis (42). Different distance measurements represent different views of the microbial community and multiple distance measurements are used to have a more comprehensive view. In order to determine the difference in community membership between BMI categories, we used the PERMANOVA test on single distance measures, with the omnibus test on the combination of all distance metrics (PermanovaG, 'GUniFrac' R package) (43).

In order to compare taxonomic abundance between groups, we used as input OTU counts. Negative binomial regression was used with BMI as a continuous variable for analysis of the microbiome while controlling for age and sex. Using multilevel modeling, the effects of confounders in study designs are examined. In this multilevel model, the study is defined as level 2 and the individual observations are level 1. At level 1, the outcome is CRC status (1=has CRC, 0=does not) and is predicted by an intercept, alpha diversity and BMI. At level 2, the level 1 regression coefficients (i.e. β_{0j}, β_{1j}, and β_{2j} for the intercept, alpha diversity and BMI regression coefficient, respectively) are modeled by the study characteristics. In this model, the level 1 regression coefficients vary among studies, which means, for example, that the effect of alpha diversity to predict CRC status varies by study and study characteristics. For precisely, we are estimating the following model:

$$logit(\mu_{ij}) = \beta_{0j} + \beta_{1j}(Alpha\ Diversity) + \beta_{2j}(BMI) + \epsilon_{ij}$$

Where the regression coefficients are modeled by study characteristics. For example, $\beta_{0j} = \gamma_{00} + \gamma_{01}(Sequencing\ Method) + \gamma_{02}(Variable\ Region) + u_{0j}$; which defines how the model parameter vary by study characteristics. In this model, the γ's represent the level two model parameters and u is the study specific error term. Estimation of this model is employed using the lme4 (linear mixed effects) package in R (44). Due to the few number of studies included in this meta-analysis, the estimation of the variance of the level 1 parameters is uncertain and should be interpreted with caution.

In order to compare taxonomic abundance between groups, we used as input OTU counts. Negative binomial regression was used with BMI as a continuous

variable (obese vs. non-obese) for analysis of the microbiome while controlling for age and sex. Multiple testing procedure was conducted on these values using BH-based false discovery rate control. The criterion to declare significance was q-value <0.2 . Furthermore, the abundance of *Bifidobacterium catenulatum* was examined among groups of obese or non-obese individuals with and without CRC. The standardized mean differences among studies was calculated using Hedge's g, a bias corrected estimate of standardized mean differences. Estimation was employed using the meta package in R (45).

Mediation analyses were also conducted. The goal of these analyses is to uncover if the relationship between BMI and CRC status is mediated by bacteria present. First, bacteria were dummy coded for presence or not for everyone. By dummy coded solely for whether an individual has a given bacteria or not, these results are not meant to show mediation among varying levels of each bacterium. Second, the relationship between BMI and CRC status was estimated by using a simple logistic regression model. Third, the classic mediation model was estimated by using the lavaan package in R (46). This model is estimated for the presence of each bacterium. Lastly, the change in the odd ratio is calculated between models. The change in the OR is an estimate of the mediation effect that a bacterium has on the relationship between BMI and CRC status.

Further exploration of whether taxonomic abundance among obese or non-obese individuals is indicative of CRC utilized random forest analyses. Random forest analysis is a machine learning/predictive modeling algorithm designed to estimate an ensemble of decision trees that are combined to give an estimate of an output. In this

study, we employed random forest analyses as a classification of obesity (obese vs. non-obese) conditional on the status of CRC. Four random forests were grown for each study dataset when possible; the forests were grown using the relative abundance of taxa with or without age and sex included at the OTU and genus level for two subsets of data that were conditioned on CRC status (CRC or adenoma). The resulting models aimed to classify individuals as obese based on the microbiota composition, and these classification models were tested with 10-fold cross validation. The receiver-operating-curve (ROC) of these classifications was also inspected for how sensitive the models are to detect obese individuals and how specific these models are to select only individuals that are obese. A measure of model quality is the area under curve (AUC), or area under the ROC, where an AUC of one is perfect prediction and an AUC of .5 is pure chance or prediction. Another benefit of using random forest analyses is that an estimate of the predictive importance of each OTU or genera is estimated. This estimate of importance is found by the predictive quality of model conditional of the ensemble trees that do not contain that specific input variable (OTU or genera in this case). All processed data and code for this analysis has been deposited at: https://github.com/GreathouseLab/CRC_BMI_meta_analysis.

Results

Database and study selection. We performed a systematic review and meta-analysis guided search of the literature. Within this search we included studies that analyzed the microbiome of the stool or tissue from patients with colon cancer, and which also had clinical information from patients on BMI. From this initial search, we identified 24 studies. After eliminating studies in which BMI information could not be obtained, 5

studies were included in the final analysis (Table 1). Given that our central hypothesis is predicated on a difference in microbial structure and composition between obese and non-obese individuals, we focused our initial analyses on the Baxter et al. study, which has adequate sample size to detect differences between these two groups (8). The remaining studies were used as comparators to support or negate any associations found between the microbiome and obesity.

Characterization of cofounders between studies. A major issue facing microbiome studies is the lack of standardized methods for collection, storage, nucleotide extraction, sequencing methodology and bioinformatic analysis. Thus, we began our analysis by characterizing the effect of 16S rRNA variable region and sequencing methods (16S rRNA or WGS) on observed OTUs and Shannon diversity on prediction of CRC. Unfortunately, we could not fully test the effect of nucleotide extraction as the Feng et al. study did not provide this information. We chose to focus on alpha diversity for this analysis given that it is a low-resolution measure, which allows for comparison across studies. Using multilevel modeling to predict CRC status we calculated the average log2 OR (logit) for each study when these level 2 predictors (variable region and sequencing method) are included in the model. The results of this analysis demonstrated that alpha diversity and obesity vary by study but do not significantly change the probability of having CRC (Figure 1A-B; Figure S1). Interestingly, the Feng et al. data set display an unusual inverse relationship between CRC ad BMI that strongly impacts prediction of CRC, possibly due to geographic and dietary differences in this population. Since all but one of the studies used the V4 16S rRNA region, it was difficult to determine if this variable had a significant impact.

Between the studies that used different extraction techniques, Zeller (GNOME DNA Isolation Kit, MP Biomedical) vs Baxter and Zackular et al.(Power Soil, Mo Bio), we did not observe an effect of extraction technique on the relationship between alpha diversity and probability of CRC (Figure 1A-B). Further, the predictive ability of 16S rRNA data, alpha diversity, to classify CRC varies among studies but using WGS does not improve this predicative ability nor does variable region choice (Figure 1A-B). Overall, among the potential confounders we tested, we did not observe a significant effect on the ability of alpha diversity to classify CRC cases and controls when controlling for obesity.

Alpha Diversity Analysis. We next sought to validate previous studies showing differences in alpha diversity between obese and non-obese individuals without CRC. In order to analyze alpha diversity within each sample study population, we calculated both richness, observed OTUs, and Shannon diversity, which considers both evenness and richness. We conducted linear modeling analysis using BMI as a continuous measurement and calculated the observed OTUs and Shannon diversity controlling for age and sex. Confirming previous microbial studies of stool from healthy (non-CRC) individuals (30), we also found significantly lower Shannon diversity in individuals that are obese without cancer from two of the 16S rRNA data sets (Baxter and Zeller et al. (WGS)) and lower richness in the Zeller et al. (16S) data; unadjusted Mann-Whitney U tests did not show this same result comparing individuals with and without obesity (Fig. 2A; Supplemental Fig S2A and Table 2). Supporting previous meta-analyses, however, studies with N<100 subjects displayed similar trends but did not reach statistical significance. When we asked if this same trend of lower Shannon

diversity was present in obese individuals with CRC, we saw no association, with the exception of the Feng dataset, which demonstrated a significantly higher alpha diversity with higher BMI both as continuous and categorical models, but not in the unadjusted analysis (Fig. 2B; Supplemental Fig S2B and Table 2). These results may be due to geography and diet of Asian populations. We chose not to analyze the *Bacteroides/Firmicutes* ratio as this has not demonstrated to be a consistent measurement of predicting obesity in human studies (30). Together, these data indicate that while there is an association between community composition and obesity in those without CRC, this association is not present in those with both obesity and CRC.

Beta Diversity Analysis. We next asked whether we could detect microbial community differences in structure between obese and normal weight individuals with or without CRC. In order to conduct this analysis, we calculated the distance matrix for each study using UniFrac or Bray-Curtis (BC) for 16S rRNA datasets, and BC or Jaccard-Sorrensen (JS) distance for WGS datasets. Further, we calculated the omnibus p-value for comparison of all distance matrices (47). In all of the data sets analyzed, except Vogtmann et al. (WGS), Zeller et al. (WGS), and Zeller et al. (16S rRNA/tissue), we observed a significant difference (omnibus p-value <0.05) in community structure between obese and non-obese individuals without CRC (Table 3; see Fig. S3A in supplemental materials). This same analysis in individuals with CRC (obese v non-obese), however, yielded only one significant observation in the Feng et al. dataset (Table 3; Supplemental Fig. S3B), supporting the observations with

community composition. Thus, similar to community composition, community structure is associated with BMI in individuals without CRC but not in those with CRC.

Taxonomic Diversity Analysis. Again, we began our taxonomic analysis comparing individuals with and without obesity among individuals without CRC as a means of validating previous studies, using as our reference the largest study dataset, Baxter et al. (8). From this analysis, controlling for age and sex, a significantly lower relative abundance of several *Ruminiococcus spp.* was identified in the two of the datasets (Zackular et al., Zeller et al. (16S rRNA stool), Zeller et al. (WGS), as well as, *Coprococcus spp.* (Baxter et al., Zackular et al., Zeller et al. (16S rRNA stool)), *Bacteroides spp.* (Baxter et al., Zackular et al., Feng et al., Vogtmannn et al., Zeller et al. (WGS)), *Bifidobacterium spp.* (Zeller et al. (WGS)) and *Akkermansia muciniphila* (Zackular et al., Zeller et al. (WGS)) (Supplemental Fig. S4 and Supplemental Table 1). When combining all differentially significant species, those from genus *Bacteroides* and *Bifidobacteria* appeared most often to differentiate individuals with and without obesity (Supplemental Table 1). While no one genera or species was found to be differentially abundant (higher or lower) between all 5 datasets comparing individuals with or without obesity among individuals without CRC, the genus *Bacteroides* contained the greatest number of differentially abundant species in individuals with obesity in all but one dataset (Supplemental Table 1).

Mediation effect of differentially abundant taxa on obesity-associated CRC classification. In order to determine if any taxa were affecting (mediating) the relationship between BMI and CRC probability, we took two approaches. The first

approach was a classical mediation test, in which we constructed three tests. First, we estimated the odds-ratio (OR) of individuals with higher BMI being more likely to be classified as having CRC. Second, we estimated the same relationship between BMI and CRC status while controlling for the mediating effect of differentially abundant bacteria. Meaning, if the bacterium mediates the relationship between BMI and CRC probability then the OR for BMI will decrease. Third, we calculated how much *change* in the OR occurred from the first to second model. Thus, from this change in ORs, we estimated how much of an effect including each taxa had on increasing or decreasing the probability of being classified as CRC for each one unit increase in BMI. From this analysis, we identified several taxa that increased or decreased the probability of CRC (Supplemental Tables 2-3). Species from the *Bacteriodes, Ruminococcus* and *Prevotella* genera, as well as, *Bifidobacterium catenulatum* decreased the probability of CRC with increasing BMI, except for two species of *Prevotella* which increased CRC probability. The mediation effect of these taxa, however, was relatively weak; less than 1% change in OR (change in probability of CRC, OR range = -9e-05 – -0.01) (Supplemental Table 2), with the majority showing a negative effect and only 8/34 showing a positive effect; none showed a significant mediation effect (Supplemental Table 2).

In our second approach, we derived an overall mediation effect using the FDR adjusted p-values (q-values) from our analysis of the Pooled BMI data (association between BMI and microbiome using all samples only, adjusting for disease status, sex and age) and Pooled DS data (association of disease status with microbiome using CRC and normal samples, adjusting for BMI, sex and age). These overall Q-values

were approximately based on 1 - (1 - q1)*(1-q2), where q1 and q2 are q values for the BMI and DS associations on the pooled data set (q value can be interpreted as the probability of being false positive, 1 - (1 - q1)*(1-q2) is the probability of being false positive in either of the associations, assuming independence between the two tests). The q-values were calculated for each data set, and q-values for taxa <20% were considered to have a significant mediating effect. Using this approach, we looked for taxa that had a significant mediating effect between studies and identified two, *Phascolarctobacterium succinatutens* and *Streptococcus salivarius*; however, they were only shared between 2/6 studies each (Supplemental Table 4). Overall, these results indicate the majority of bacteria associated with CRC and BMI decrease the odds of CRC in individuals with obesity, but only weakly.

In addition, to determine if previously identified CRC-associated taxa, *F. nucleatum, F. prausnitzii, B. fragilis,* or *A. muciniphila,* were altered in individuals with obesity in their ability to differentiate CRC from non-CRC, we calculated the log2 odds ratios for each species. (Supplementary Fig. S5). Overall, among individuals with obesity, *F. nucleatum* consistently showed stronger prediction (log2 OR) of CRC.

Ability of the microbiome to classify obesity-associated CRC. Given that previous studies have demonstrated the predicative capability of the microbiome in generating classifiers for CRC, we next asked whether a taxonomic consortium could accurately predict obesity-associated CRC. Using the machine learning method random forest, we calculated importance scores among obese individuals at the OTU or genus level using 10-fold cross-validation in individuals with adenomas or CRC. These values were then used to calculate area under the receiver operator curve using age and sex

as co-variates or the microbiome alone. Among all obese individuals, the average of all AUC values predicting CRC cases at the OTU and genus level was 0.66 (0.47-0.84) and 0.68 (0.47-0.94), respectively (Fig. 3B). Similarly, among obese adenoma cases, average AUC values at the OTU and genus level were 0.61 (0.48-0.86) and 0.60 (0.52-0.73), respectively (Fig. 3A); demonstrating high heterogeneity among studies in predicting CRC or adenomas in obese individuals. Lastly, we sought to validate CRC classifiers developed by Baxter et al. and Zeller et al. by agnostic application of our random forest classifier on each dataset using all genera or OTUs. While Zeller et al. used a more complex statistical approach to construct their classifier, we choose to apply the same method (48) to each study for the purposes of comparison, which was almost identical to Baxter et al. (excluding smoking and hemoglobin test results). Overall, the microbiome by itself or controlling for BMI, age and sex, had low and variable AUC values (OTU; AUC=0.53-0.79; Genus; AUC=0.59-0.81) in most studies. We were able, however, to validate the classifier from the Baxter et al. and Feng et al. studies; our AUC values were 0.79 (Baxter et al.) and 0.81 as compared to Baxter et al. (AUC=0.84) and Feng et al. (AUC=0.96). Although we could not approach the classifier values from the Zeller et al. study (AUC=0.84; without FOBT), this was likely due to the difference in their approach in building the classifier. In general, these data indicate that the microbiome together with clinical data, and likely FOBT or similar tests, could have diagnostic utility.

Analysis of Inferred Taxonomic Function. Multiple studies have demonstrated that taxonomic abundance alone does not accurately reflect the metabolic function of the entire community. Thus, we interrogated the metabolic potential of the bacterial

community using the bioinformatics tool PICRUSt in order to obtain predicted functions. Among individuals without CRC, we identified biotin synthesis and biotin metabolism inversely correlated with BMI in the Vogtmannn (WGS) stool samples, and the urea cycle (M0029) inversely correlated with BMI in both Vogtmannn (WGS) and Feng (WGS) (Figure 4A-B). However, none of these predicted functions differentiated obese individuals among all studies. When we conducted this analysis in individuals with CRC, no shared correlations were identified when comparing obese and non-obese. Predicted functional analysis therefore, did not further distinguish obesity-associated CRC from those with CRC and normal BMI.

Discussion.

Evidence clearly demonstrates an intimate link between inflammation, obesity, and the microbiome (34, 36, 49-56). *In vivo,* multiple studies indicate an interaction or mediating effect of the microbiome in promoting colon tumorigenesis in the presence of a high-fat diet or genetic-induced obesity (21, 22, 57-59). In this study, using BMI as a measure of obesity, we were able to initiate the first analysis addressing this outstanding question in human subjects.

This is the most comprehensive high-resolution study of the microbiome in individuals with and without obesity among those with CRC, using multiple sequencing platforms and methods. In this meta-analysis, we describe both obesity- and CRC-associated results. First, we found both community structure and composition in stool and tissue samples from individuals with CRC are independent of BMI. Second, we identified a weak effect of the majority of species associated with both BMI and CRC on risk of CRC. Lastly, we show the microbiome, by itself or modeled with age and

sex, is insufficient to classify adenomas or CRC from obese controls. However, when controlling for clinical variables and BMI, we are able to achieve similar levels of CRC classification to other studies (48). Overall, by combining species-level resolution from 16S rRNA and WGS data, we were able to define the microbial community structure and function at a high resolution, revealing overall a weak effect of the microbiome on mediating CRC risk among individuals with obesity as compared to those with normal BMI.

While this study did not identify any strong universal BMI-associated microbial biomarkers of CRC, many mechanisms are likely key in driving the increased risk of CRC in obese individuals that we could not account for in this study. These include tumor location (left vs right), mutation profile, differentiation, mismatch repair status, and diet; some of which have shown to differentiate individuals with obesity among those with CRC (60-63). A high fat diet may be more important than BMI or obesity in driving the deleterious changes in the microbiome in individuals with obesity. In support, feeding a high-fat diet to K-ras^{G12Dint} mice is sufficient to drive tumorigenesis from 30% to 60% (33). Moreover, when feces from high fat fed mice (K-ras^{G12Dint}) are transferred to healthy (K-ras^{G12Dint}) mice, tumor burden is increased along with diminished immune cell recruitment (33). This was prevented, however, when supplemented with butyrate, which also increased *Bifidobacterium* abundance as compared to mice not supplemented with butyrate (33). We also found several species of *Bifidobacterium* lower in individuals with obesity among those with and without CRC. Interestingly, butyrate and butyrate producing bacteria were shown to be increased in African-American men after switching to a traditional high fiber, low-fat

rural African diet (64). Again, similar to the results of high-fat feeding promoting CRC, which was abrogated with butyrate treatment, the aforementioned study found that the high-fat Western diet of African-Americans was associated with higher secondary bile acids, known promoters of carcinogenesis. Together, these studies indicate that a high fat diet, specifically from saturated fats, may be interacting with the microbiome to create a pro-inflammatory environment conducive to colon carcinogenesis.

Other possible mechanisms explaining the increased risk of CRC in individuals with obesity include a lack of balance in key immune regulatory cells, specifically regulatory T cells (Tregs) and B lymphocytes. Demonstration that Tregs are important in promoting colon tumorigenesis, indicates that species that can control their activation may be important in controlling CRC development. Specifically, when *B. pseudocatenulatum* (CECT 7765) was given orally to obese mice, it increased Tregs and reduced pro-inflammatory cytokines (IL-17A and TNF-a), which further supports the hypothesis that certain species may protect against chronic inflammation and development of CRC (51). In reports measuring dietary inflammatory factors (empirical dietary inflammatory pattern), individuals with higher inflammatory scores had fewer tumor-associated adaptive anti-tumor immune cells suggesting immune evasion (65). Moreover, using this same approach, higher inflammatory scores were associated with higher tumor-associated *F. nucleatum* in CRC (66). Again, these findings support a distinct influence of diet on the microbiome in CRC development apart from obesity.

BMI is crude measure of obesity, and other more accurate measures (e.g. waist circumference, adipokines, etc.) are required to fully explore the relationship between obesity, inflammation, and the microbiome in development of CRC. An exemplar of

this relationship is demonstrated for lung cancer, wherein the use of BMI demonstrates a lower risk of lung cancer is associated with higher BMI but use of waist circumference or waist to hip ratio demonstrates and increased risk of lung cancer (67). Thus, this study sets the stage for future research to consider adding measures of adiposity beyond BMI when studying the etiology and risk of CRC, as well as, other cancers influenced by obesity.

This study has several strengths, as well as, limitations. One important strength, was the ability to use multiple peer-reviewed studies that had similar study designs and sequencing methods. As the Microbiome Quality Control Project demonstrated, multiple factors (e.g. DNA extraction method) can contribute to differential findings between studies, and thus our ability to control for these confounding factors reduced this bias (68-71). Also, the ability to confirm the presence of multiple taxa using separate sequencing methods, 16S rRNA and WGS methods, further strengthened the design of this analysis. The limitations of this study include small sample sizes in the majority of studies, use of only one anthropometric measurement of obesity and lack of other informative factors including dietary fat intake, previous weight loss prior to CRC diagnosis, microbial metabolites and biofilm presence. Sample size is a key limitation when looking at the relationship between obesity (BMI) and the microbiome as previously demonstrated (30). Only the Baxter (8) study was sufficiently powered to detect a significant difference in the microbial community between normal and obese individuals. While we were able to identify taxa that differentiated obese and normal individuals with CRC in this study specifically, these taxa were not consistent across all studies, indicating that other factors such as metabolites, biofilm or the immune

system are stronger contributors to this relationship. While we were able to derive inferred function from the microbial sequences, without more intensive direct measurement of the metabolites (i.e. mass spectrometry), we cannot fully assess these differences. Additionally, studies have illustrated this lack of relationship between specific taxa and CRC, and instead identified a stronger association with the presence of biofilm formation. Lastly, animal studies demonstrating that CRC promotion by a high fat diet was independent of obesity supports our findings and suggests that dietary fat has a greater impact than obesity on the microbiome and its tumor-promoting capacity in CRC etiology. This will therefore be important to consider in the obesity-CRC relationship in future research.

Overall, our validation of microbiome-based classifiers indicates this approach, in combination with FOBT or FIT tests, is well supported for continued development. More important, these data along with other studies indicate that diet, rather than obesity, is creating a pro-inflammatory microbial community increasing CRC risk. Hence, characterizing the role of the diet in addition to the microbiome in CRC etiology is necessary, which will require more detailed molecular analyses and well-designed longitudinal human studies to identify early stage dietary and microbial biomarkers prior to disease.

Funding Sources

This work was supported by the Baylor University Summer Sabbatical Grant (PI – summer salary support). Nick Chia was supported by the NCI award R01CA179243.

Competing interests

James White is a significant shareholder in the company Resphera Insight Inc. All other authors declare that they have no competing interests

Author contributions

KLG conceived of the study, analysis plan, analyzed data, interpreted results and participated in writing and review; JW downloaded and processed all sequencing data; JC, GDJ and NP conducted statistical analyses; BGP processed data; JW, KLG, JC, GDJ, NP, BGP, and NC provided technical and data interpretation assistance and manuscript review. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank the participants in each of the studies for their time dedication to supporting colon cancer research. Also, we thank all of the authors from each study for making their data publicly available for analysis, and to allow for transparency and reproducibly. Our thanks to the members of the Baylor Writing Group, Joe Taube, Karen Melton, Elise King, Elyssia Gallagher, for their feedback and editing.

References

1. **Arnold M, Pandeya N, Byrnes G, Renehan PAG, Stevens GA, Ezzati PM, Ferlay J, Miranda JJ, Romieu I, Dikshit R, Forman D, Soerjomataram I.** 2015. Global burden of cancer attributable to high body-mass index in 2012: a population-based study. Lancet Oncol **16:**36-46.
2. **Whiteman DC, Wilson LF.** 2016. The fractions of cancer attributable to modifiable factors: A global review. Cancer Epidemiol **44:**203-221.

3. **Giovannucci E, Ascherio A, Rimm EB, Colditz GA, Stampfer MJ, Willett WC.** 1995. Physical activity, obesity, and risk for colon cancer and adenoma in men. Ann Intern Med **122:**327-334.
4. **Engeland A, Tretli S, Austad G, Bjorge T.** 2005. Height and body mass index in relation to colorectal and gallbladder cancer in two million Norwegian men and women. Cancer Causes Control **16:**987-996.
5. **Calle EE, Kaaks R.** 2004. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. Nat Rev Cancer **4:**579-591.
6. **Flegal KM, Kruszon-Moran D, Carroll MD, Fryar CD, Ogden CL.** 2016. Trends in Obesity Among Adults in the United States, 2005 to 2014. JAMA **315:**2284-2291.
7. **Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, Zhang D, Xia H, Xu X, Jie Z, Su L, Li X, Li X, Li J, Xiao L, Huber-Schonauer U, Niederseer D, Xu X, Al-Aama JY, Yang H, Wang J, Kristiansen K, Arumugam M, Tilg H, Datz C, Wang J.** 2015. Gut microbiome development along the colorectal adenoma-carcinoma sequence. Nat Commun **6:**6528.
8. **Baxter NT, Ruffin MTt, Rogers MA, Schloss PD.** 2016. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. Genome Med **8:**37.
9. **Nakatsu G, Li X, Zhou H, Sheng J, Wong SH, Wu WK, Ng SC, Tsoi H, Dong Y, Zhang N, He Y, Kang Q, Cao L, Wang K, Zhang J, Liang Q, Yu J, Sung JJ.** 2015. Gut mucosal microbiome across stages of colorectal carcinogenesis. Nat Commun **6:**8727.
10. **Chung L, Thiele Orberg E, Geis AL, Chan JL, Fu K, DeStefano Shields CE, Dejea CM, Fathi P, Chen J, Finard BB, Tam AJ, McAllister F, Fan H, Wu X, Ganguly S, Lebid A, Metz P, Van Meerbeke SW, Huso DL, Wick EC, Pardoll DM, Wan F, Wu S, Sears CL, Housseau F.** 2018. Bacteroides fragilis Toxin Coordinates a Pro-carcinogenic Inflammatory Cascade via Targeting of Colonic Epithelial Cells. Cell Host Microbe **23:**203-214 e205.
11. **Dejea CM, Fathi P, Craig JM, Boleij A, Taddese R, Geis AL, Wu X, DeStefano Shields CE, Hechenbleikner EM, Huso DL, Anders RA, Giardiello FM, Wick EC, Wang H, Wu S, Pardoll DM, Housseau F, Sears CL.** 2018. Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. Science **359:**592-597.
12. **Purcell RV, Pearson J, Aitchison A, Dixon L, Frizelle FA, Keenan JI.** 2017. Colonization with enterotoxigenic Bacteroides fragilis is associated with early-stage colorectal neoplasia. PLoS One **12:**e0171602.
13. **Bullman S, Pedamallu CS, Sicinska E, Clancy TE, Zhang X, Cai D, Neuberg D, Huang K, Guevara F, Nelson T, Chipashvili O, Hagan T, Walker M, Ramachandran A, Diosdado B, Serna G, Mulet N, Landolfi S, Ramon YCS, Fasani R, Aguirre AJ, Ng K, Elez E, Ogino S, Tabernero J, Fuchs CS, Hahn WC, Nuciforo P, Meyerson M.** 2017. Analysis of Fusobacterium persistence and antibiotic response in colorectal cancer. Science **358:**1443-1448.
14. **Yang Y, Weng W, Peng J, Hong L, Yang L, Toiyama Y, Gao R, Liu M, Yin M, Pan C, Li H, Guo B, Zhu Q, Wei Q, Moyer MP, Wang P, Cai S, Goel A, Qin H, Ma Y.** 2017. Fusobacterium nucleatum Increases Proliferation of Colorectal

Cancer Cells and Tumor Development in Mice by Activating Toll-Like Receptor 4 Signaling to Nuclear Factor-kappaB, and Up-regulating Expression of MicroRNA-21. Gastroenterology **152:**851-866 e824.

15. **Strauss J, Kaplan GG, Beck PL, Rioux K, Panaccione R, Devinney R, Lynch T, Allen-Vercoe E.** 2011. Invasive potential of gut mucosa-derived Fusobacterium nucleatum positively correlates with IBD status of the host. Inflamm Bowel Dis **17:**1971-1978.
16. **Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, Clancy TE, Chung DC, Lochhead P, Hold GL, El-Omar EM, Brenner D, Fuchs CS, Meyerson M, Garrett WS.** 2013. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. Cell Host Microbe **14:**207-215.
17. **Tahara T, Yamamoto E, Suzuki H, Maruyama R, Chung W, Garriga J, Jelinek J, Yamano HO, Sugai T, An B, Shureiqi I, Toyota M, Kondo Y, Estecio MR, Issa JP.** 2014. Fusobacterium in colonic flora and molecular features of colorectal carcinoma. Cancer Res **74:**1311-1318.
18. **Zackular JP, Baxter NT, Iverson KD, Sadler WD, Petrosino JF, Chen GY, Schloss PD.** 2013. The gut microbiome modulates colon tumorigenesis. MBio **4:**e00692-00613.
19. **Everard A, Lazarevic V, Gaia N, Johansson M, Stahlman M, Backhed F, Delzenne NM, Schrenzel J, Francois P, Cani PD.** 2014. Microbiome of prebiotic-treated mice reveals novel targets involved in host response during obesity. ISME J **8:**2116-2130.
20. **Ley RE, Turnbaugh PJ, Klein S, Gordon JI.** 2006. Microbial ecology: human gut microbes associated with obesity. Nature **444:**1022-1023.
21. **Pfalzer AC, Kamanu FK, Parnell LD, Tai AK, Liu Z, Mason JB, Crott JW.** 2016. Interactions between the colonic transcriptome, metabolome, and microbiome in mouse models of obesity-induced intestinal cancer. Physiol Genomics **48:**545-553.
22. **Pfalzer AC, Nesbeth PD, Parnell LD, Iyer LK, Liu Z, Kane AV, Chen CY, Tai AK, Bowman TA, Obin MS, Mason JB, Greenberg AS, Choi SW, Selhub J, Paul L, Crott JW.** 2015. Diet- and Genetically-Induced Obesity Differentially Affect the Fecal Microbiome and Metabolome in Apc1638N Mice. PLoS One **10:**e0135758.
23. **Turnbaugh PJ, Backhed F, Fulton L, Gordon JI.** 2008. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. Cell Host Microbe **3:**213-223.
24. **Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI.** 2009. A core gut microbiome in obese and lean twins. Nature **457:**480-484.
25. **Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI.** 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature **444:**1027-1031.
26. **Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, Griffin NW, Lombard V, Henrissat B, Bain JR, Muehlbauer MJ, Ilkayeva O, Semenkovich**

CF, Funai K, Hayashi DK, Lyle BJ, Martini MC, Ursell LK, Clemente JC, Van Treuren W, Walters WA, Knight R, Newgard CB, Heath AC, Gordon JI. 2013. Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science **341:**1241214.

27. **Bervoets L, Van Hoorenbeeck K, Kortleven I, Van Noten C, Hens N, Vael C, Goossens H, Desager KN, Vankerckhoven V.** 2013. Differences in gut microbiota composition between obese and lean children: a cross-sectional study. Gut Pathog **5:**10.
28. **Ferrer M, Ruiz A, Lanza F, Haange SB, Oberbach A, Till H, Bargiela R, Campoy C, Segura MT, Richter M, von Bergen M, Seifert J, Suarez A.** 2013. Microbiota from the distal guts of lean and obese adolescents exhibit partial functional redundancy besides clear differences in community structure. Environ Microbiol **15:**211-226.
29. **Walters WA, Xu Z, Knight R.** 2014. Meta-analyses of human gut microbes associated with obesity and IBD. FEBS Lett **588:**4223-4233.
30. **Sze MA, Schloss PD.** 2016. Looking for a Signal in the Noise: Revisiting Obesity and the Microbiome. MBio **7**.
31. **Caer C, Rouault C, Le Roy T, Poitou C, Aron-Wisnewsky J, Torcivia A, Bichet JC, Clement K, Guerre-Millo M, Andre S.** 2017. Immune cell-derived cytokines contribute to obesity-related inflammation, fibrogenesis and metabolic deregulation in human adipose tissue. Sci Rep **7:**3000.
32. **Chehimi M, Vidal H, Eljaafari A.** 2017. Pathogenic Role of IL-17-Producing Immune Cells in Obesity, and Related Inflammatory Diseases. J Clin Med **6**.
33. **Schulz MD, Atay C, Heringer J, Romrig FK, Schwitalla S, Aydin B, Ziegler PK, Varga J, Reindl W, Pommerenke C, Salinas-Riester G, Bock A, Alpert C, Blaut M, Polson SC, Brandl L, Kirchner T, Greten FR, Polson SW, Arkan MC.** 2014. High-fat-diet-mediated dysbiosis promotes intestinal carcinogenesis independently of obesity. Nature **514:**508-512.
34. **Everard A, Belzer C, Geurts L, Ouwerkerk JP, Druart C, Bindels LB, Guiot Y, Derrien M, Muccioli GG, Delzenne NM, de Vos WM, Cani PD.** 2013. Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. Proc Natl Acad Sci U S A **110:**9066-9071.
35. **Dao MC, Everard A, Aron-Wisnewsky J, Sokolovska N, Prifti E, Verger EO, Kayser BD, Levenez F, Chilloux J, Hoyles L, Consortium MI-O, Dumas ME, Rizkalla SW, Dore J, Cani PD, Clement K.** 2016. Akkermansia muciniphila and improved metabolic health during a dietary intervention in obesity: relationship with gut microbiome richness and ecology. Gut **65:**426-436.
36. **Greer RL, Dong X, Moraes AC, Zielke RA, Fernandes GR, Peremyslova E, Vasquez-Perez S, Schoenborn AA, Gomes EP, Pereira AC, Ferreira SR, Yao M, Fuss IJ, Strober W, Sikora AE, Taylor GA, Gulati AS, Morgun A, Shulzhenko N.** 2016. Akkermansia muciniphila mediates negative effects of IFNgamma on glucose metabolism. Nat Commun **7:**13329.
37. **Drewes JL, White JR, Dejea CM, Fathi P, Iyadorai T, Vadivelu J, Roslani AC, Wick EC, Mongodin EF, Loke MF, Thulasi K, Gan HM, Goh KL, Chong HY, Kumar S, Wanyiri JW, Sears CL.** 2017. High-resolution bacterial 16S rRNA

gene profile meta-analysis and biofilm status reveal common colorectal cancer consortia. NPJ Biofilms Microbiomes **3:**34.

38. **Daquigan N, Seekatz AM, Greathouse KL, Young VB, White JR.** 2017. High-resolution profiling of the gut microbiome reveals the extent of Clostridium difficile burden. NPJ Biofilms Microbiomes **3:**35.
39. **Guerrero-Preston R, White JR, Godoy-Vitorino F, Rodriguez-Hilario A, Navarro K, Gonzalez H, Michailidi C, Jedlicka A, Canapp S, Bondy J, Dziedzic A, Mora-Lagos B, Rivera-Alvarez G, Ili-Gangas C, Brebi-Mieville P, Westra W, Koch W, Kang H, Marchionni L, Kim Y, Sidransky D.** 2017. High-resolution microbiome profiling uncovers Fusobacterium nucleatum, Lactobacillus gasseri/johnsonii, and Lactobacillus vaginalis associated to oral and oropharyngeal cancer in saliva from HPV positive and HPV negative patients treated with surgery and chemo-radiation. Oncotarget **8:**110931-110948.
40. **Byrd AL, Perez-Rogers JF, Manimaran S, Castro-Nallar E, Toma I, McCaffrey T, Siegel M, Benson G, Crandall KA, Johnson WE.** 2014. Clinical PathoScope: rapid alignment and filtration for accurate pathogen identification in clinical samples using unassembled sequencing data. BMC Bioinformatics **15:**262.
41. **Oh J, Byrd AL, Deming C, Conlan S, Program NCS, Kong HH, Segre JA.** 2014. Biogeography and individuality shape function in the human skin metagenome. Nature **514:**59-64.
42. **Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Dore J, Meta HITC, Antolin M, Artiguenave F, Blottiere HM, Almeida M, Brechot C, Cara C, Chervaux C, Cultrone A, Delorme C, Denariaz G, et al.** 2011. Enterotypes of the human gut microbiome. Nature **473:**174-180.
43. **Chen J, Bittinger K, Charlson ES, Hoffmann C, Lewis J, Wu GD, Collman RG, Bushman FD, Li H.** 2012. Associating microbiome composition with environmental covariates using generalized UniFrac distances. Bioinformatics **28:**2106-2113.
44. **Bates D, Mächler M, Bolker B, Walker S.** 2015. Fitting Linear Mixed-Effects Models Using lme4. 2015 **67:**48.
45. **Schwarzer G.** 2007. meta: An R package for meta-analysis, R News, R News. https://cran.r-project.org/doc/Rnews/Rnews_2007-3.pdf.
46. **Rosseel Y.** 2012. lavaan: An R Package for Structural Equation Modeling. 2012 **48:**36.
47. **Chen J, King E, Deek R, Wei Z, Yu Y, Grill D, Ballman K, Stegle O.** 2018. An omnibus test for differential distribution analysis of microbiome sequencing data. Bioinformatics **34:**643-651.
48. **Sze MA, Schloss PD.** 2018. Leveraging Existing 16S rRNA Gene Surveys To Identify Reproducible Biomarkers in Individuals with Colorectal Tumors. MBio **9**.

49. **Kellermayer R, Dowd SE, Harris RA, Balasa A, Schaible TD, Wolcott RD, Tatevian N, Szigeti R, Li Z, Versalovic J, Smith CW.** 2011. Colonic mucosal DNA methylation, immune response, and microbiome patterns in Toll-like receptor 2-knockout mice. FASEB J **25:**1449-1460.
50. **Garidou L, Pomie C, Klopp P, Waget A, Charpentier J, Aloulou M, Giry A, Serino M, Stenman L, Lahtinen S, Dray C, Iacovoni JS, Courtney M, Collet X, Amar J, Servant F, Lelouvier B, Valet P, Eberl G, Fazilleau N, Douin-Echinard V, Heymes C, Burcelin R.** 2015. The Gut Microbiota Regulates Intestinal CD4 T Cells Expressing RORgammat and Controls Metabolic Disease. Cell Metab **22:**100-112.
51. **Moya-Perez A, Neef A, Sanz Y.** 2015. Bifidobacterium pseudocatenulatum CECT 7765 Reduces Obesity-Associated Inflammation by Restoring the Lymphocyte-Macrophage Balance and Gut Microbiota Structure in High-Fat Diet-Fed Mice. PLoS One **10:**e0126976.
52. **Schneeberger M, Everard A, Gomez-Valades AG, Matamoros S, Ramirez S, Delzenne NM, Gomis R, Claret M, Cani PD.** 2015. Akkermansia muciniphila inversely correlates with the onset of inflammation, altered adipose tissue metabolism and metabolic disorders during obesity in mice. Sci Rep **5:**16643.
53. **Zeng H, Ishaq SL, Zhao FQ, Wright AG.** 2016. Colonic inflammation accompanies an increase of beta-catenin signaling and Lachnospiraceae/Streptococcaceae bacteria in the hind gut of high-fat diet-fed mice. J Nutr Biochem **35:**30-36.
54. **Zhao S, Liu W, Wang J, Shi J, Sun Y, Wang W, Ning G, Liu R, Hong J.** 2017. Akkermansia muciniphila improves metabolic profiles by reducing inflammation in chow diet-fed mice. J Mol Endocrinol **58:**1-14.
55. **Thaiss CA, Levy M, Grosheva I, Zheng D, Soffer E, Blacher E, Braverman S, Tengeler AC, Barak O, Elazar M, Ben-Zeev R, Lehavi-Regev D, Katz MN, Pevsner-Fischer M, Gertler A, Halpern Z, Harmelin A, Aamar S, Serradas P, Grosfeld A, Shapiro H, Geiger B, Elinav E.** 2018. Hyperglycemia drives intestinal barrier dysfunction and risk for enteric infection. Science **359:**1376-1383.
56. **Liu R, Hong J, Xu X, Feng Q, Zhang D, Gu Y, Shi J, Zhao S, Liu W, Wang X, Xia H, Liu Z, Cui B, Liang P, Xi L, Jin J, Ying X, Wang X, Zhao X, Li W, Jia H, Lan Z, Li F, Wang R, Sun Y, Yang M, Shen Y, Jie Z, Li J, Chen X, Zhong H, Xie H, Zhang Y, Gu W, Deng X, Shen B, Xu X, Yang H, Xu G, Bi Y, Lai S, Wang J, Qi L, Madsen L, Wang J, Ning G, Kristiansen K, Wang W.** 2017. Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention. Nat Med **23:**859-868.
57. **Higashimura Y, Naito Y, Takagi T, Uchiyama K, Mizushima K, Ushiroda C, Ohnogi H, Kudo Y, Yasui M, Inui S, Hisada T, Honda A, Matsuzaki Y, Yoshikawa T.** 2016. Protective effect of agaro-oligosaccharides on gut dysbiosis and colon tumorigenesis in high-fat diet-fed mice. Am J Physiol Gastrointest Liver Physiol **310:**G367-375.
58. **Liu W, Crott JW, Lyu L, Pfalzer AC, Li J, Choi SW, Yang Y, Mason JB, Liu Z.** 2016. Diet- and Genetically-induced Obesity Produces Alterations in the

Microbiome, Inflammation and Wnt Pathway in the Intestine of Apc(+/1638N) Mice: Comparisons and Contrasts. J Cancer **7:**1780-1790.

59. **Koh GY, Kane A, Lee K, Xu Q, Wu X, Roper J, Mason JB, Crott JW.** 2018. Parabacteroides distasonis attenuates toll-like receptor 4 signaling and Akt activation and blocks colon tumor formation in high-fat diet-fed azoxymethane-treated mice. Int J Cancer doi:10.1002/ijc.31559.
60. **Hanyuda A, Cao Y, Hamada T, Nowak JA, Qian ZR, Masugi Y, da Silva A, Liu L, Kosumi K, Soong TR, Jhun I, Wu K, Zhang X, Song M, Meyerhardt JA, Chan AT, Fuchs CS, Giovannucci EL, Ogino S, Nishihara R.** 2017. Body mass index and risk of colorectal carcinoma subtypes classified by tumor differentiation status. Eur J Epidemiol **32:**393-407.
61. **Li Liua FKT, Xuehong Zhanga, Jonathan A. Nowaka, Zhi Rong Qian, Tsuyoshi Hamada, Daniel Nevo, Susan Bullman, Kosuke Mima, Keisuke Kosumi, Annacarolina da Silva, Mingyang Song, Yin Cao, Tyler S. Twombly, Yan Shi, Hongli Liu, Mancang Gu, Hideo Koh, Wanwan Li, Chunxia Du, Yang Chen, Chenxi Li, Wenbin Li, Raaj S. Mehta, Kana Wu, Molin Wang, Aleksander D. Kostic, Marios Giannakis, Wendy S. Garrett, Curtis Hutthenhower, Andrew T. Chan, Charles S. Fuchs, Reiko Nishiharab, Shuji Ogino'Correspondence information about the author Shuji Oginob,Email the author Shuji Ogino, Edward L. Giovannuccib.** 2018. Diets That Promote Colon Inflammation Associate With Risk of Colorectal Carcinomas That Contain Fusobacterium nucleatum. Clinical Gastroenterology and Hepatology doi:10.1016/j.cgh.2018.04.030.
62. **Hanyuda A, Ogino S, Qian ZR, Nishihara R, Song M, Mima K, Inamura K, Masugi Y, Wu K, Meyerhardt JA, Chan AT, Fuchs CS, Giovannucci EL, Cao Y.** 2016. Body mass index and risk of colorectal cancer according to tumor lymphocytic infiltrate. Int J Cancer **139:**854-868.
63. **Sinicrope FA, Foster NR, Yoon HH, Smyrk TC, Kim GP, Allegra CJ, Yothers G, Nikcevich DA, Sargent DJ.** 2012. Association of obesity with DNA mismatch repair status and clinical outcome in patients with stage II or III colon carcinoma participating in NCCTG and NSABP adjuvant chemotherapy trials. J Clin Oncol **30:**406-412.
64. **O'Keefe SJ, Li JV, Lahti L, Ou J, Carbonero F, Mohammed K, Posma JM, Kinross J, Wahl E, Ruder E, Vipperla K, Naidoo V, Mtshali L, Tims S, Puylaert PG, DeLany J, Krasinskas A, Benefiel AC, Kaseb HO, Newton K, Nicholson JK, de Vos WM, Gaskins HR, Zoetendal EG.** 2015. Fat, fibre and cancer risk in African Americans and rural Africans. Nat Commun **6:**6342.
65. **Liu L, Nishihara R, Qian ZR, Tabung FK, Nevo D, Zhang X, Song M, Cao Y, Mima K, Masugi Y, Shi Y, da Silva A, Twombly T, Gu M, Li W, Hamada T, Kosumi K, Inamura K, Nowak JA, Drew DA, Lochhead P, Nosho K, Wu K, Wang M, Garrett WS, Chan AT, Fuchs CS, Giovannucci EL, Ogino S.** 2017. Association Between Inflammatory Diet Pattern and Risk of Colorectal Carcinoma Subtypes Classified by Immune Responses to Tumor. Gastroenterology **153:**1517-1530.e1514.
66. **Liu L, Tabung FK, Zhang X, Nowak JA, Qian ZR, Hamada T, Nevo D, Bullman S, Mima K, Kosumi K, da Silva A, Song M, Cao Y, Twombly TS, Shi**

Y, Liu H, Gu M, Koh H, Li W, Du C, Chen Y, Li C, Li W, Mehta RS, Wu K, Wang M, Kostic AD, Giannakis M, Garrett WS, Hutthenhower C, Chan AT, Fuchs CS, Nishihara R, Ogino S, Giovannucci EL. Diets That Promote Colon Inflammation Associate With Risk of Colorectal Carcinomas That Contain Fusobacterium nucleatum. Clinical Gastroenterology and Hepatology doi:10.1016/j.cgh.2018.04.030.

67. **Yu D, Zheng W, Johansson M, Lan Q, Park Y, White E, Matthews CE, Sawada N, Gao YT, Robien K, Sinha R, Langhammer A, Kaaks R, Giovannucci EL, Liao LM, Xiang YB, Lazovich D, Peters U, Zhang X, Bueno-de-Mesquita B, Willett WC, Tsugane S, Takata Y, Smith-Warner SA, Blot W, Shu XO.** 2018. Overall and Central Obesity and Risk of Lung Cancer: A Pooled Analysis. J Natl Cancer Inst doi:10.1093/jnci/djx286.

68. **Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, Schwager E, Crabtree J, Ma S, Microbiome Quality Control Project C, Abnet CC, Knight R, White O, Huttenhower C.** 2017. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. Nat Biotechnol **35:**1077-1086.

69. **Sinha R, Abnet CC, White O, Knight R, Huttenhower C.** 2015. The microbiome quality control project: baseline study design and future directions. Genome Biol **16:**276.

70. **Olson ND, Morrow JB.** 2012. DNA extract characterization process for microbial detection methods development and validation. BMC Res Notes **5:**668.

71. **Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, Gonzalez A, Kosciolek T, McCall LI, McDonald D, Melnik AV, Morton JT, Navas J, Quinn RA, Sanders JG, Swafford AD, Thompson LR, Tripathi A, Xu ZZ, Zaneveld JR, Zhu Q, Caporaso JG, Dorrestein PC.** 2018. Best practices for analysing microbiomes. Nat Rev Microbiol doi:10.1038/s41579-018-0029-9.

Figure Legends

Figure 1:
Variance in ability of alpha diversity to predict odds (log2) of CRC controlling for obesity and study confounders. The Log_2 odds ratio of CRC using observed OTUs (left panel) or Shannon Index (right panel) as predictors. The multilevel model includes obesity (level 1), and sequencing method (16S rRNA or WGS) and variable region (V4 or V3-4) (level 2) as coefficients.

Figure 2:
Alpha diversity in individuals with or without obesity and with or without CRC
A) Observed OTUs and Shannon diversity in individuals without CRC B) or with CRC comparing individuals with or without obesity. Reporting p-values are from Mann-Whitney U Test comparing the alpha diversity of individuals with or without obesity.

Figure 3:
Microbial classifiers of CRC and obesity-associated CRC.
A) Receiver Operating Curve (ROC) for the random forest classification analyses for obese vs. non-obese in individuals with CRC for each study. AUC is the 10-fold cross validated area under the curve. B) ROC for the random forest classification analyses of obese vs. non-obese in individuals with adenomas for each study. Due to a lack of cases with adenomas in some studies a random forest was not possible and are therefore not shown. C) ROC for the random forest classification analyses of CRC vs non-CRC in each dataset adjusted for BMI, age and sex.

Figure 4:
Pathway abundance analysis in individuals with or without obesity among individuals with or without CRC.
Relative abundance of KEGG metabolic pathways (16S rRNA) or modules (WGS) inferred from PICRUSt or HUMAnN, respectively. Significance was calculated using the Wilcox test correction for multiple hypothesis testing; asterisks are representative of significance at adjusted p-value <0.2.

Supplemental Figures

Figure S1:
Probability of having CRC using alpha diversity as a predictor among individuals with obesity. Predicted probability of having CRC using A) observed OTUs or B) Shannon Index.

Figure S2:
Alpha diversity by BMI in individuals with or without CRC
A) Observed OTUs and Shannon diversity in individuals without CRC comparing BMI and alpha diversity metric, observed OTUs or Shannon diversity respectively. B)

Observed OTUs and Shannon diversity in individuals with CRC comparing BMI and alpha diversity metric, observed OTUs or Shannon diversity respectively.

Figure S3:
Beta diversity in induvial with or without obesity and with or without CRC
A) Differences in community composition between individuals with and without obesity among those without CRC B) or with CRC. The axes were found using PCoA using Bray-Curtis distances among points with the proportion of variance accounted for by each axis reported. Points are colored by obesity status.

Figure S4:
Differential abundance of taxa associated with obesity and CRC taxa.
OTUs (16S rRNA) or species (WGS) log 10 scale relative abundance of *Ruminiococcus spp., Coprococcus spp., Bacteroides spp., Bifidobacterium spp*. and *Akkermansia muciniphila*. P-values were calculated using negative binomial regression using abundance as a count and including age and sex as covariates. Significant differences between obese v non-obese with or without CRC are denoted by an asterisk (FDR adjusted p-value <0.1)

Figure S5:
Ability of CRC-associated taxa to predict CRC among individuals with obesity.
For each species identified from previous CRC microbiome studies, *F. nucleatum, F. prausnitzii, B. fragilis,* or *A. muciniphila,* the $\log_2$ odd ratio was calculated for individuals with obesity to determine odds of being classified as having CRC.

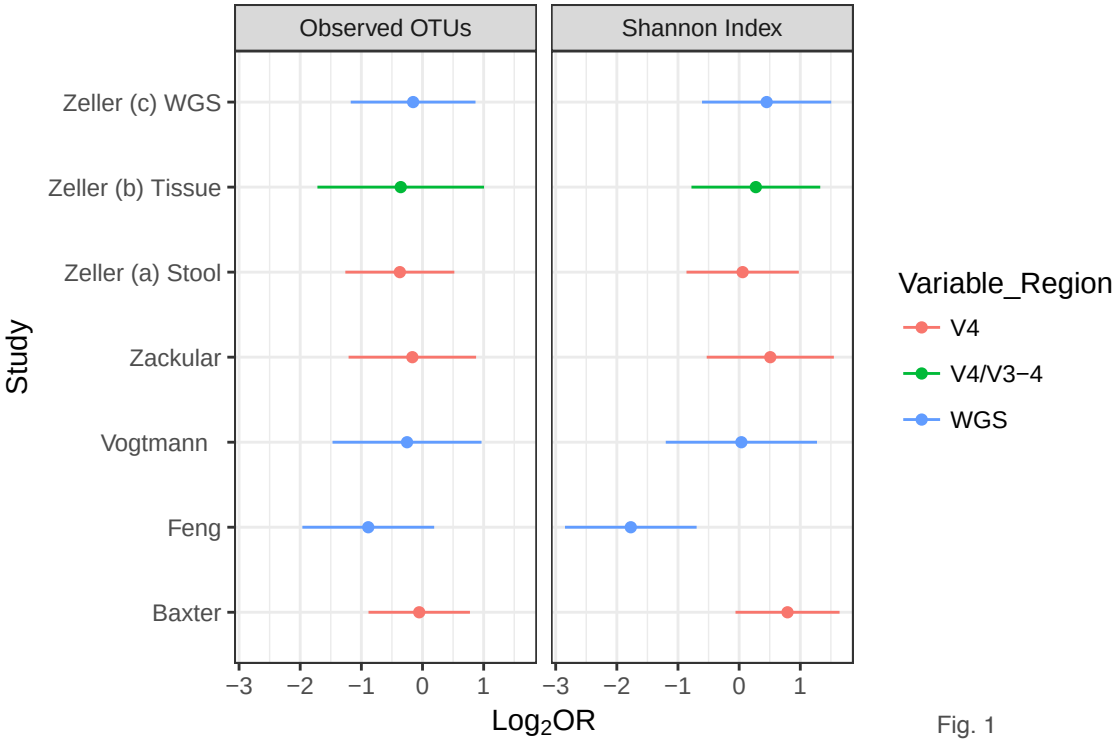

Observed OTUs
Shannon Index
Zeller (c) WGS
Zeller (b) Tissue
Zeller (a) Stool
Zackular
Vogtmann
Feng
Baxter
Study
−3
−2
−1
0
1
Log2OR
Variable_Region
V4
V4/V3−4
WGS

Fig. 1

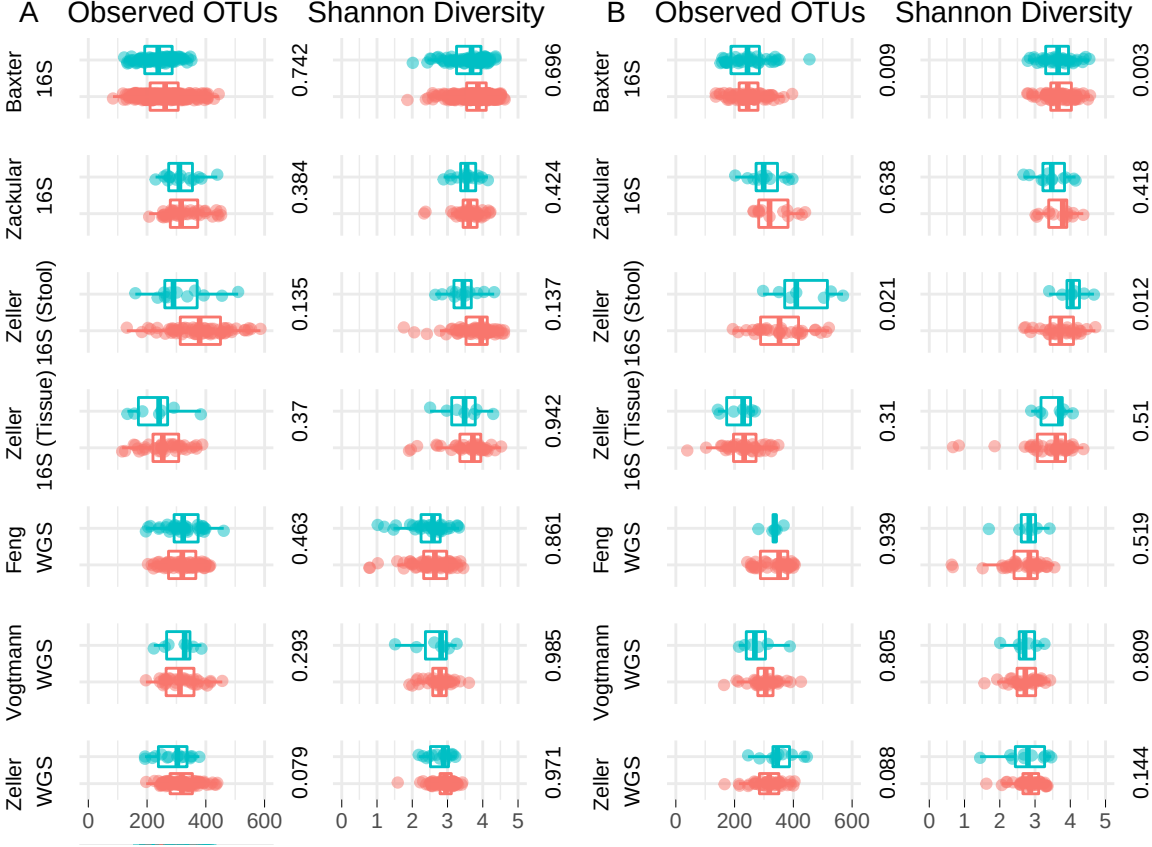
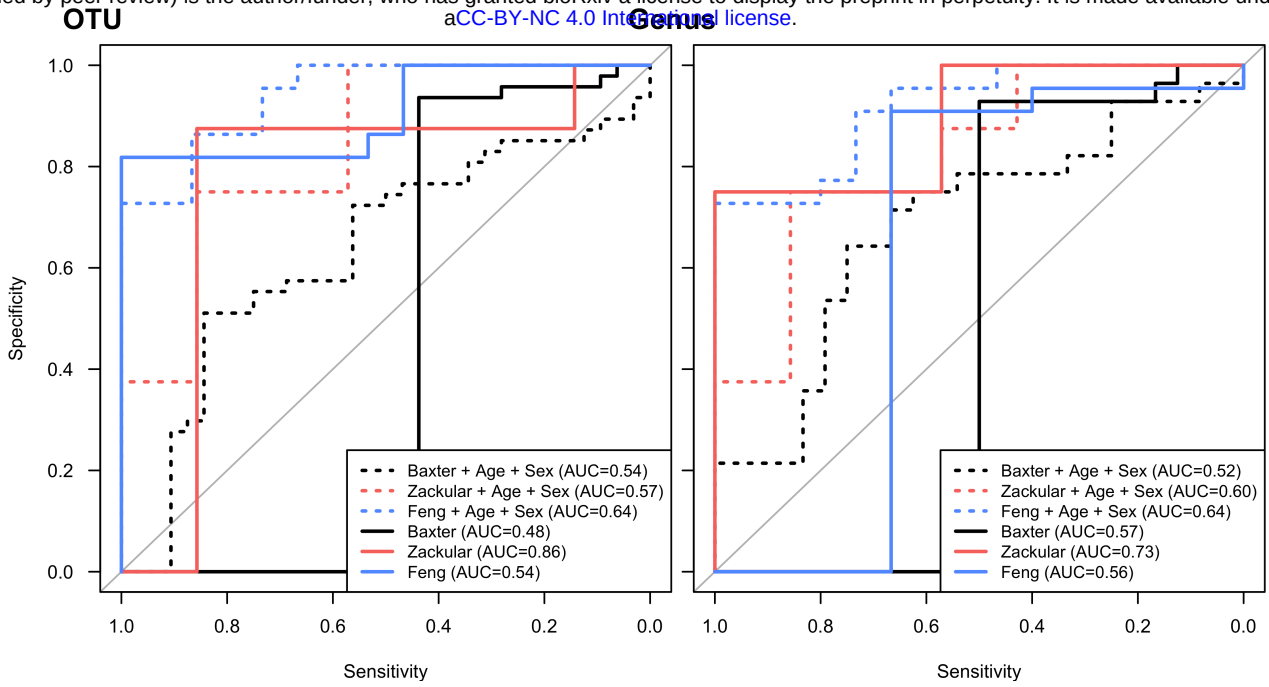
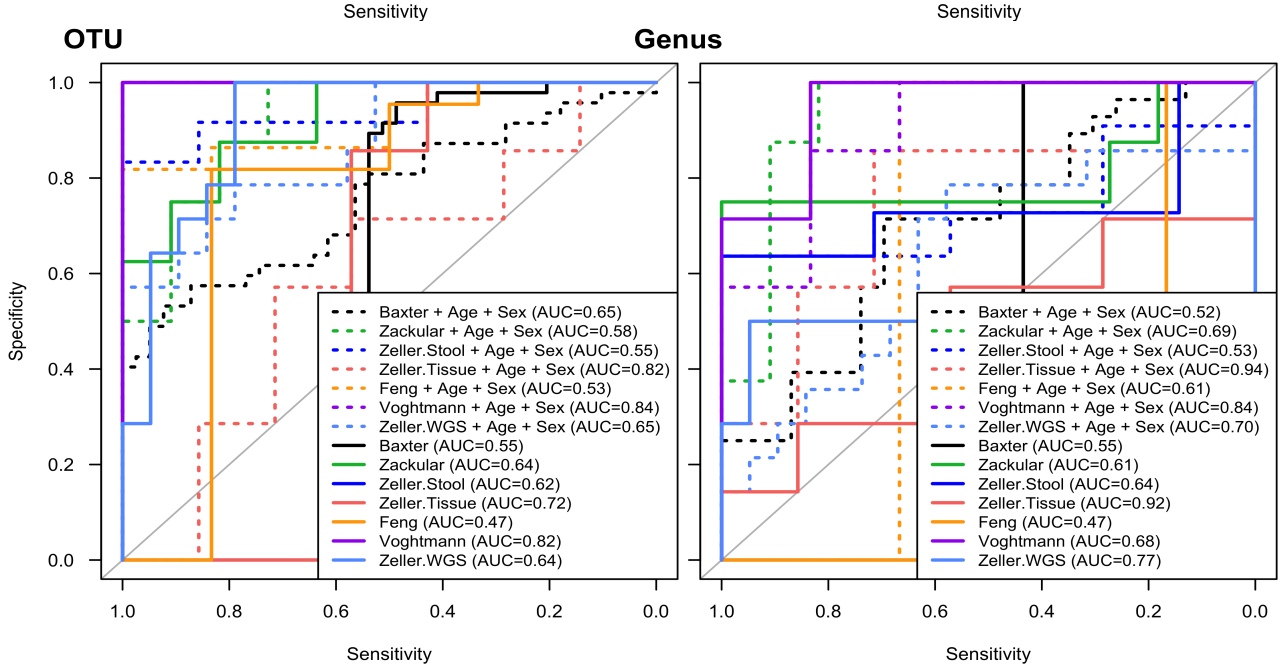

A
Observed OTUs
Shannon Diversity
B
Observed OTUs
Shannon Diversity
Baxter 16S
Zackular 16S
Zeller 16S (Stool)
Zeller 16S (Tissue)
Feng WGS
Vogtmann WGS
Zeller WGS
0.742
0.696
0.384
0.424
0.135
0.137
0.37
0.942
0.463
0.861
0.293
0.985
0.079
0.971
0.009
0.003
0.638
0.418
0.021
0.012
0.31
0.51
0.939
0.519
0.805
0.809
0.088
0.144
0
200
400
600
0
1
2
3
4
5
BMI_Cat
Non−Obese
Obese

Fig. 2

A



B



C

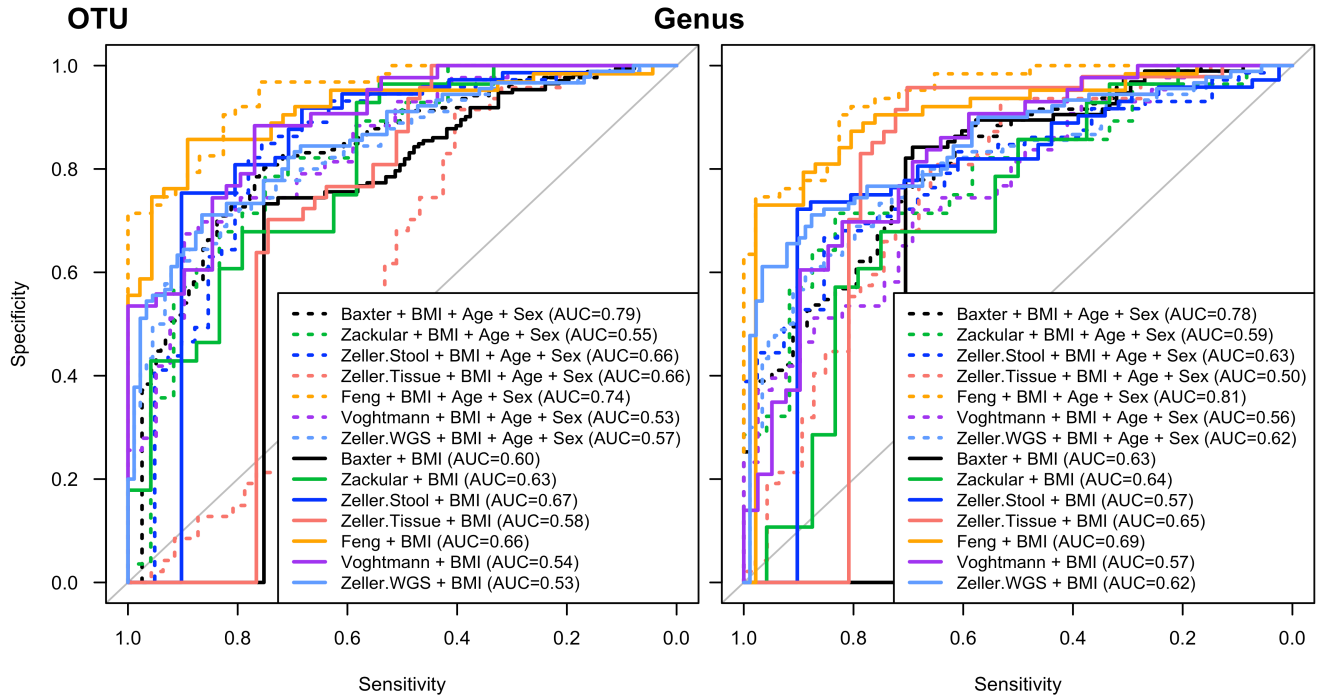


Fig. 3

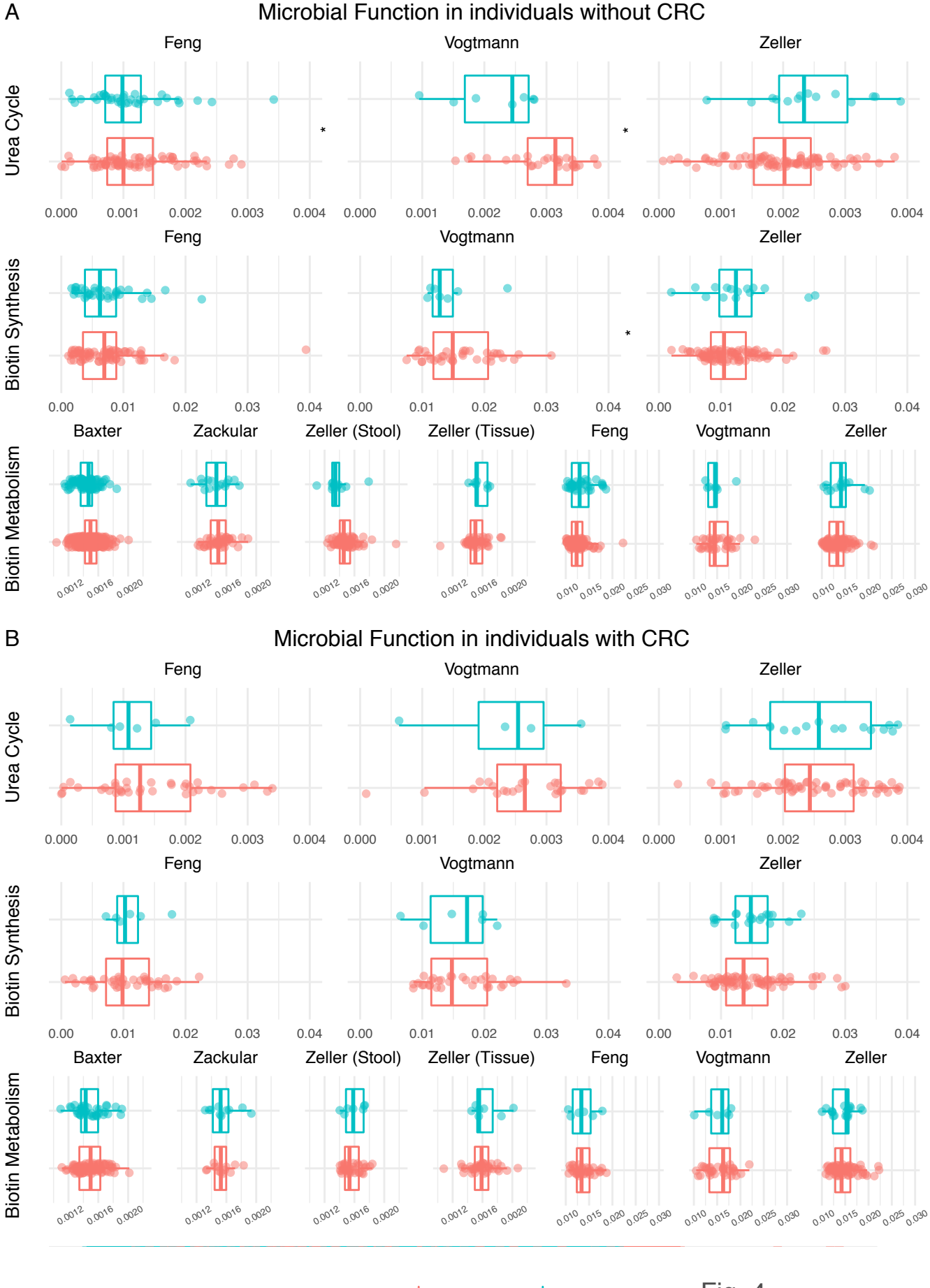

A
Microbial Function in individuals without CRC
Feng
Vogtmann
Zeller
Urea Cycle
Biotin Synthesis
Baxter
Zackular
Zeller (Stool)
Zeller (Tissue)
Biotin Metabolism
B
Microbial Function in individuals with CRC
BMI_Cat
Non−Obese
Obese

Fig. 4

Table 1: Summary of obesity, demographic, sequencing, and reads for included data sources

						Sample Size by Group		Average BMI		
Database (Study)	**Sample Type**	**Sequencing Method**	**Primers ***	**Species/OTUs[#]**	**Sample Source (N)**	**Obese**	**Non-Obese**	**Obese**	**Non-Obese**	***P* (obese v non-obese)^**
Baxter et al. (2016)	Stool	16S rRNA	V4	9997	carcinoma(318); control (172)	118	368	34.05	24.93	< .001
Feng et al. (2015)	Stool	WGS	NA	408772	adenoma (42); carcinoma (41); control (55)	43	113	31.98	25.62	< .001
Vogtmann et al. (2016)	Stool	WGS	NA	356748	carcinoma (52); control (52)	13	69	32.57	23.58	< .001
Zackular et al. (2014)	Stool	16S rRNA	V4	24990	adenoma (30); carcinoma (30); control (30)	26	56	33.84	25.09	0.002
Zeller et al. (2014)a	Stool	16S rRNA	V4	9969	control (75); CRC (41)	19	108	32.53	24.13	< .001
Zeller et al. (2014)b	Tissue	16S rRNA	V4/V3-4	9988	carcinoma (48); carcinoma-adjacent (48)	14	80	33.3	24.5	< .001
Zeller et al. (2014)c	Stool	WGS	NA	327491	adenoma (42); carcinoma (53); control (297)	34	160	32.15	24.18	< .001

*1 Primers are NA for WGS

Average number of observed taxanomonic units (16S rRNA) or species (WGS). Averages were rounded to nearest whole number.

^ Test of the equality of proportion of individuals that are obese vs. non-obese.

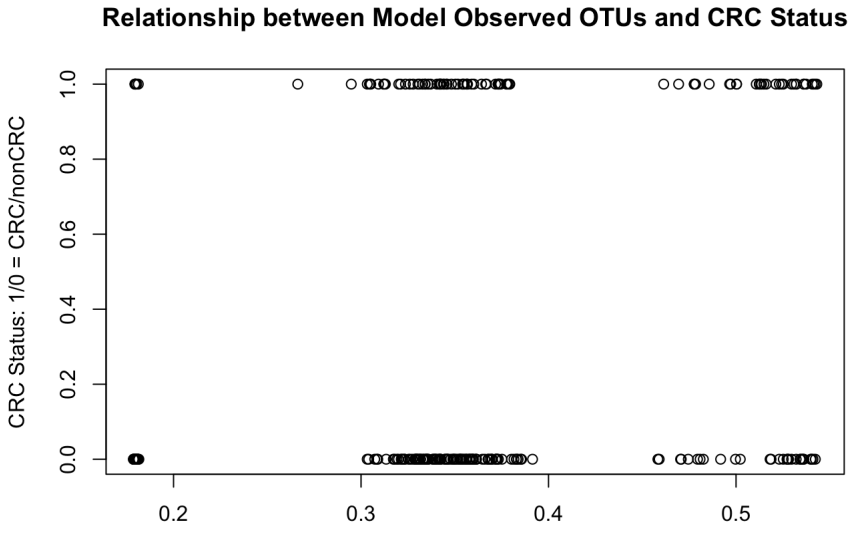
Table 2: Summary of Alpha-Diversity Analysis

Type	Reference (Sample)	Observed		Shannon	
		Linear (Est.)	p-value	Linear (Est.)	p-value
16S rRNA	**Baxter et al. (Stool)**				
	Non-CRC - Ob v Non-Ob	-22.756	0.064	**-0.173**	**0.046**
	CRC - Ob v Non-Ob	-1.283	0.930	-0.059	0.564
	Zackular et al. (Stool)				
	Non-CRC - Ob v Non-Ob	-36.737	0.248	-0.240	0.217
	CRC - Ob v Non-Ob	7.006	0.811	0.053	0.794
	Zeller et al. (Stool)				
	Non-CRC - Ob v Non-Ob	**-86.434**	**0.049**	-0.455	0.051
	CRC - Ob v Non-Ob	-42.696	0.551	-0.020	0.953
	Zeller et al. (Tissue)				
	Non-CRC - Ob v Non-Ob	-14.373	0.682	-0.019	0.949
	CRC - Ob v Non-Ob	-12.389	0.683	0.140	0.690
WGS	**Vogtmann et al. (Stool)**				
	Non-CRC - Ob v Non-Ob	-25.023	0.399	-0.165	0.372
	CRC - Ob v Non-Ob	-9.187	0.698	-0.025	8.971
	Feng et al. (Stool)				
	Non-CRC - Ob v Non-Ob	-27.211	0.204	-0.327	0.085
	CRC - Ob v Non-Ob	21.947	0.348	0.389	0.183
	Zeller et al. (Stool)				
	Non-CRC - Ob v Non-Ob	-21.877	0.157	**-0.219**	**0.041**
	CRC - Ob v Non-Ob	3.592	0.832	-0.031	0.813

Table 3: Summary of Beta-Diversity Analysis

Type	Reference (Sample)	UniFrac		GUniFrac		WUniFrac		Bray-Curtis		Jaccard		Omnibus test
		F	p-value	*F*	p-value	*F*	p-value	*F*	p-value	*F*	p-value	p-value
16S rRNA	**Baxter et al. (Stool)**											
	Non-CRC - Ob v Non-Ob	**1.852**	**0.011**	**1.871**	**0.027**	1.798	0.111	**1.792**	**0.019**			**0.024**
	CRC - Ob v Non-Ob	1.156	0.175	1.221	0.179	0.897	0.422	1.452	0.069			0.136
	Zackular et al. (Stool)											
	Non-CRC - Ob v Non-Ob	1.314	0.088	**1.784**	**0.022**	**2.824**	**0.007**	**1.494**	**0.054**			**0.015**
	CRC - Ob v Non-Ob	1.191	0.166	1.479	0.093	2.183	0.050	1.368	0.094			0.117
	Zeller et al. (Stool)											
	Non-CRC - Ob v Non-Ob	**1.691**	**0.006**	**1.565**	**0.040**	1.387	0.180	1.393	0.101			**0.014**
	CRC - Ob v Non-Ob	1.132	0.208	1.127	0.280	1.394	0.178	1.142	0.276			0.327
	Zeller et al. (Tissue)											
	Non-CRC - Ob v Non-Ob	1.159	0.189	1.411	0.060	1.596	0.095	**1.597**	**0.027**			0.066
	CRC - Ob v Non-Ob	1.353	0.056	1.125	0.275	0.880	0.548	1.106	0.307			0.136
WGS	**Vogtmann et al. (Stool)**											
	Non-CRC - Ob v Non-Ob							0.897	0.485	1.041	0.422	0.508
	CRC - Ob v Non-Ob							1.086	0.357	1.068	0.355	0.426
	Feng et al. (Stool)											
	Non-CRC - Ob v Non-Ob							**3.129**	**0.003**	**4.144**	**0.008**	**0.004**
	CRC - Ob v Non-Ob							**2.923**	**0.019**	**4.972**	**0.008**	**0.010**
	Zeller et al. (Stool)											
	Non-CRC - Ob v Non-Ob							1.821	0.081	**2.892**	**0.047**	0.065
	CRC - Ob v Non-Ob							1.025	0.382	0.972	0.424	0.461

A


Relationship between Model Observed OTUs and CRC Status
CRC Status: 1/0 = CRC/nonCRC
0.0
0.2
0.4
0.6
0.8
1.0
0.2
0.3
0.4
0.5

B

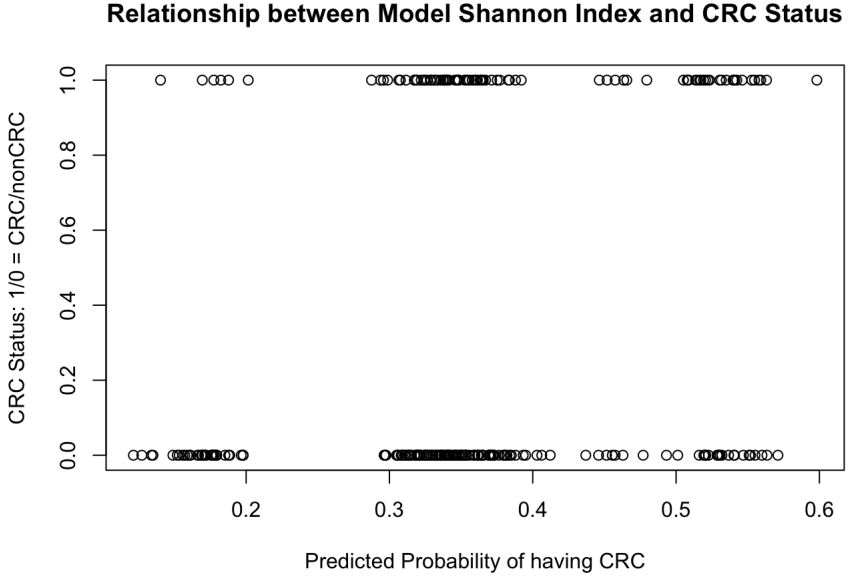

Relationship between Model Shannon Index and CRC Status
CRC Status: 1/0 = CRC/nonCRC
0.0
0.2
0.4
0.6
0.8
1.0
0.2
0.3
0.4
0.5
0.6
Predicted Probability of having CRC

Fig. S1

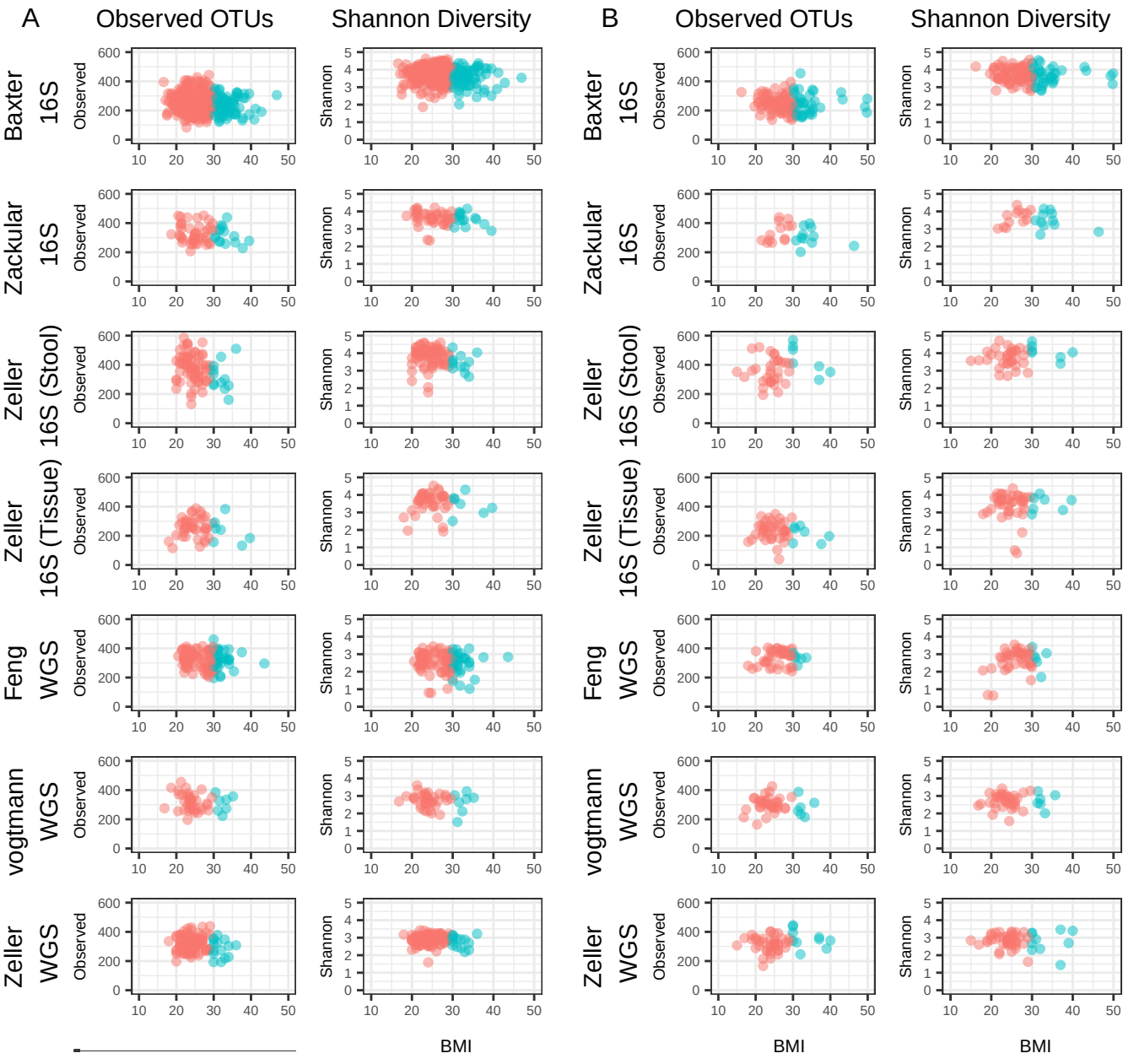

A
Observed OTUs
Shannon Diversity
B
Observed OTUs
Shannon Diversity
Baxter 16S
Zackular 16S
Zeller 16S (Stool)
Zeller 16S (Tissue)
Feng WGS
vogtmann WGS
Zeller WGS
Observed
Shannon
BMI
BMI_Cat
Non−Obese
Obese

Fig. S2

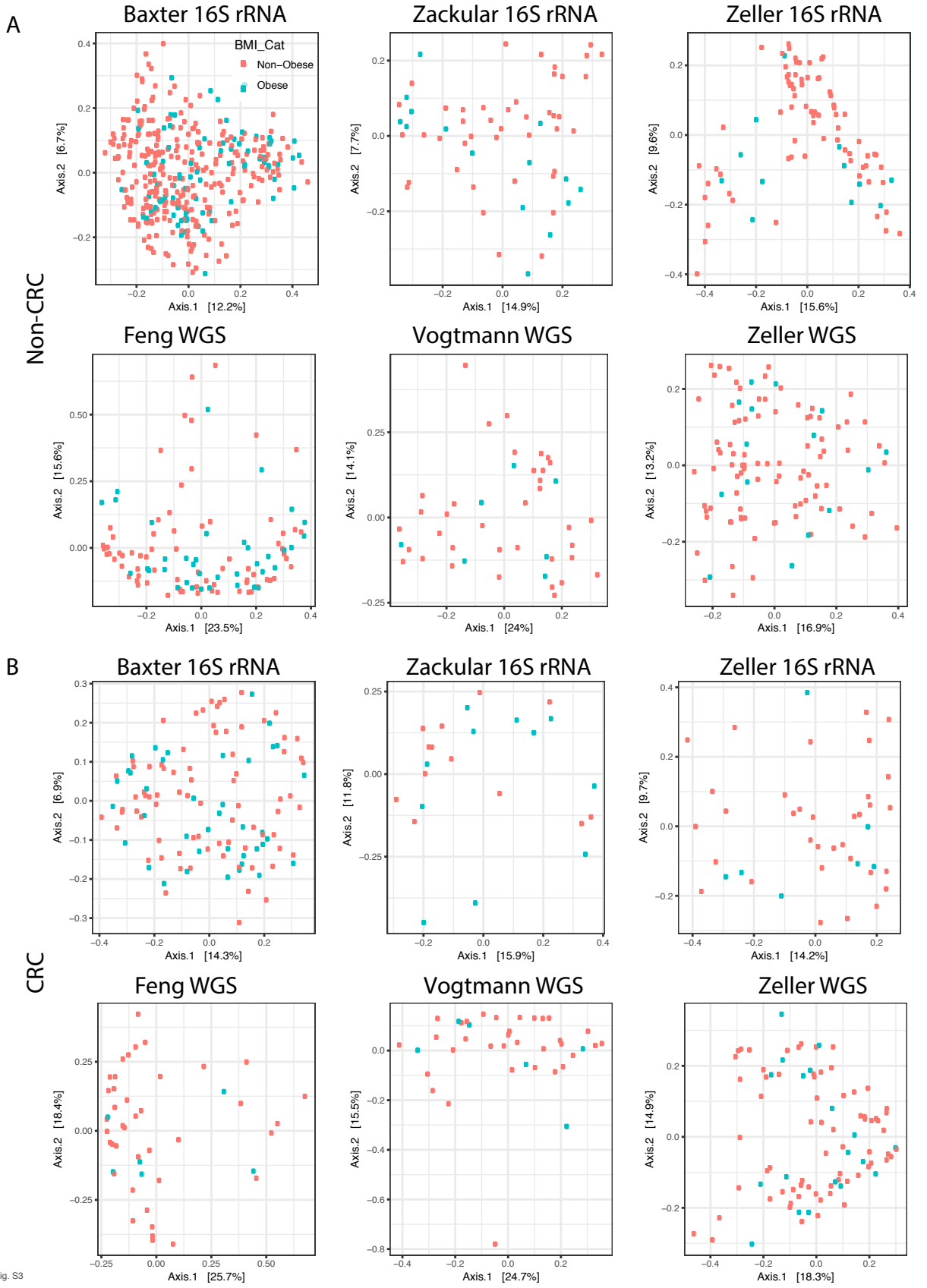
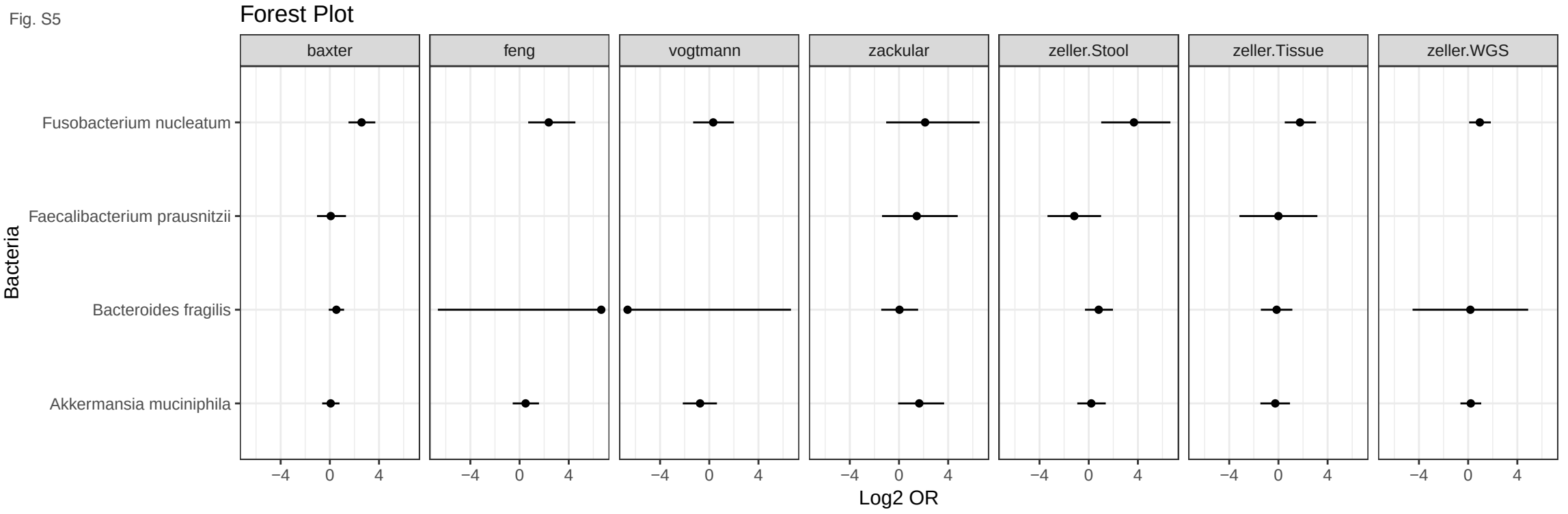

A
Non-CRC
Baxter 16S rRNA
Zackular 16S rRNA
Zeller 16S rRNA
Feng WGS
Vogtmann WGS
Zeller WGS
BMI_Cat
Non−Obese
Obese
Axis.1 [12.2%]
Axis.2 [6.7%]
Axis.1 [14.9%]
Axis.2 [7.7%]
Axis.1 [15.6%]
Axis.2 [9.6%]
Axis.1 [23.5%]
Axis.2 [15.6%]
Axis.1 [24%]
Axis.2 [14.1%]
Axis.1 [16.9%]
Axis.2 [13.2%]
B
CRC
Baxter 16S rRNA
Zackular 16S rRNA
Zeller 16S rRNA
Feng WGS
Vogtmann WGS
Zeller WGS
Axis.1 [14.3%]
Axis.2 [6.9%]
Axis.1 [15.9%]
Axis.2 [11.8%]
Axis.1 [14.2%]
Axis.2 [9.7%]
Axis.1 [25.7%]
Axis.2 [18.4%]
Axis.1 [24.7%]
Axis.2 [15.5%]
Axis.1 [18.3%]
Axis.2 [14.9%]

Fig. S3


Akkermansia muciniphila
Ruminococcus spp.
Bacteroides spp.
Bifidobacterium spp.
Coprococcus spp.
16S
WGS
Baxter
Zackular
Zeller (Stool)
Zeller (Tissue)
Feng
Vogtmann
Zeller
nonCRC CRC
0.001
0.100
*
**
BMI_Cat
Non−Obese
Obese
Status_Cat
nonCRC
CRC

Fig. S4

Fig. S5


Forest Plot
baxter
feng
vogtmann
zackular
zeller.Stool
zeller.Tissue
zeller.WGS
Fusobacterium nucleatum
Faecalibacterium prausnitzii
Bacteroides fragilis
Akkermansia muciniphila
Bacteria
−4
0
4
Log2 OR

Supplemental Table 1: Summary of Species Level Differential Abundance

Genus	Reference (Sample)	Non-CRC Obese v Non-Obese					Genus	Reference (Sample)	CRC - Ob v Non-Ob				
		Pvalue	Qvalue	BMI	2.50%	97.50%			Pvalue	Qvalue	BMI	2.50%	97.50%
	Species							*Species*					
Akkermansia	**Baxter et al. (Stool)**						**Akkermansia**	**Baxter et al. (Stool)**					
	NONE							OTUAkkermansia_muciniphila:Verrucomicrobia;Akkermansia	0.008	0.090	-0.069	-0.120	-0.018
	Zackular et al. (Stool)							**Zackular et al. (Stool)**					
	OTUAkkermansia_muciniphila:Verrucomicrobia;Akkermansia	0.001	0.004	-0.240	-0.378	-0.101		NONE					
	Zeller et al. (Stool)							**Zeller et al. (Stool)**					
	NONE							OTUotu3490:Akkermansia_muciniphila:Verrucomicrobia;Akkermansia	0.000	0.004	-0.104	-0.159	-0.050
	Feng et al. (Stool)							**Feng et al. (Stool)**					
	NONE							Akkermansia muciniphila	0.024	0.115	0.183	0.024	0.342
	Vogtmannn et al. (Stool)							**Vogtmannn et al. (Stool)**					
Ruminococcus	**Baxter et al. (Stool)**						**Ruminococcus**	**Baxter et al. (Stool)**					
	OTURuminococcus_lactaris:Firmicutes;Ruminococcus	0.001	0.019	0.060	0.024	0.096		NONE					
	Zackular et al. (Stool)							**Zackular et al. (Stool)**					
	OTUotu1316:Ruminococcus_flavefaciens:Firmicutes;Ruminococcus	0.000	0.000	-0.151	-0.171	-0.131		OTURuminococcus_callidus:Firmicutes;Ruminococcus	0.000	0.000	-0.110	-0.129	-0.091
	OTUotu1473:Clostridium_methylpentosum:Ruminococcus_albus:Firmicutes;unassigned	0.000	0.002	-0.453	-0.700	-0.205		OTUotu1072:Ruminococcus_callidus:Ruminococcus_champanellensis:Ruminococcus_flavefaciens:Firmicutes;Ruminococcus	0.000	0.000	-0.071	-0.098	-0.044
	OTUotu148:Ruminococcus_bromii:Firmicutes;Ruminococcus	0.008	0.033	-0.287	-0.499	-0.075		OTURuminococcus_champanellensis:Firmicutes;Ruminococcus	0.002	0.010	-0.233	-0.381	-0.085
								OTUotu1794:Ruminococcus_flavefaciens:Firmicutes;Ruminococcus	0.013	0.050	-0.807	-1.446	-0.169
								OTUotu1473:Clostridium_methylpentosum:Ruminococcus_albus:Firmicutes;unassigned	0.000	0.002	0.337	0.156	0.519
	Zeller et al. (Stool)							**Zeller et al. (Stool)**					
	OTURuminococcus_champanellensis:Firmicutes;Ruminococcus	0.000	0.007	-0.234	-0.358	-0.109		OTURuminococcus_callidus:Firmicutes;Ruminococcus	0.028	0.199	0.175	0.019	0.331
	OTUotu1488:Ruminococcus_bromii:Firmicutes;Ruminococcus	0.003	0.042	-0.151	-0.249	-0.052							
	OTUotu4879:Blautia_faecis:Clostridium_boliviensis:Clostridium_celerecrescens:Clostridium_hathewayi:Clostridium_saccharolyticum:Clostridium_sphenoides:Ruminococcus_lactaris:Firmicutes;unassigned	0.003	0.043	-0.227	-0.377								
	OTUotu1:Ruminococcus_bromii:Firmicutes;Ruminococcus	0.004	0.051	-0.419	-0.701	-0.137							
	OTUotu6413:Ruminococcus_champanellensis:Firmicutes;Ruminococcus	0.005	0.064	-0.197	-0.335	-0.059							
	Ruminococcus albus	0.017	0.169	-0.053	-0.096	-0.010							
	Ruminococcus flavefaciens	0.035	0.192	-0.050	-0.096	-0.004							
	Feng et al. (Stool)							**Feng et al. (Stool)**					
	NONE							NONE					
	Vogtmannn et al. (Stool)							**Vogtmannn et al. (Stool)**					
	Ruminococcus sp. 5_1_39BFAA	0.860	0.983	0.004	-0.042	0.050		[Ruminococcus] gnavus	0.308	0.675	0.028	-0.026	0.083
								Ruminococcus flavefaciens	0.364	0.713	-0.036	-0.115	0.042
Bifidobacterium	**Baxter et al. (Stool)**						**Bifidobacterium**	**Baxter et al. (Stool)**					
	NONE							NONE					
	Zackular et al. (Stool)							**Zackular et al. (Stool)**					
	NONE							OTUBifidobacterium_adolescentis:Bifidobacterium_stercoris:Actinobacteria;I	0.000	0.000	0.025	0.019	0.032
								OTUBifidobacterium_bifidum:Actinobacteria;Bifidobacterium	0.000	0.000	0.048	0.030	0.066
	OTUBifidobacterium_catenulatum:Bifidobacterium_kashiwanohense:Bifidobacterium_pseudocatenulatum:Actinobacteria;Bifidobacterium	0.656	0.769	0.054	-0.184	0.292		OTUBifidobacterium_catenulatum:Bifidobacterium_kashiwanohense:Bifidobacterium_pseudocatenulatum:Actinobacteria;Bifidobacterium	0.000	0.002	-0.033	-0.051	-0.015
	Zeller et al. (Stool)							**Zeller et al. (Stool)**					
	OTUBifidobacterium_catenulatum:Bifidobacterium_kashiwanohense:Bifidobacterium_pseudocatenulatum:Actinobacteria;Bifidobacterium	0.008	0.082	-0.272	-0.472	-0.072		Bifidobacterium catenulatum	0.008	0.096	-0.119	-0.207	-0.032
	Bifidobacterium catenulatum	0.005	0.088	-0.166	-0.281	-0.051		Bifidobacterium pseudocatenulatum	0.012	0.129	-0.095	-0.170	-0.021
	Bifidobacterium adolescentis	0.028	0.192	-0.140	-0.265	-0.015							
	Feng et al. (Stool)							**Feng et al. (Stool)**					
	Bifidobacterium dentium	0.021	0.067	0.144	0.022	0.266		Bifidobacterium catenulatum	0.001	0.005	-0.298	-0.466	-0.130
								Bifidobacterium longum	0.037	0.135	0.112	0.007	0.217
	Vogtmannn et al. (Stool)							**Vogtmannn et al. (Stool)**					
	Bifidobacterium adolescentis	0.877	0.985	0.014	-0.163	0.191		Bifidobacterium adolescentis	0.305	0.675	0.100	-0.091	0.291
	Bifidobacterium bifidum	0.218	0.581	0.068	-0.040	0.177		Bifidobacterium longum	0.311	0.675	-0.056	-0.163	0.052
	Bifidobacterium longum	0.200	0.581	0.075	-0.040	0.190							
	Bifidobacterium catenulatum	0.009	0.145	0.211	0.052	0.370							
Bacteroides	**Baxter et al. (Stool)**						**Bacteroides**	**Baxter et al. (Stool)**					
	OTUBacteroides_eggerthii:Bacteroidetes;Bacteroides	0.009	0.099	-0.171	-0.300	-0.042		OTUBacteroides_plebeius:Bacteroidetes;Bacteroides	0.000	0.000	0.037	0.023	0.051
	OTUBacteroides_fluxus:Bacteroidetes;Bacteroides	0.000	0.000	0.049	0.036	0.063		OTUBacteroides_finegoldii:Bacteroidetes;Bacteroides	0.002	0.045	-0.164	-0.267	-0.061
	Zackular et al. (Stool)							**Zackular et al. (Stool)**					
	OTUBacteroides_salyersiae:Bacteroidetes;Bacteroides	0.000	0.000	0.147	0.114	0.180		OTUBacteroides_nordii:Bacteroidetes;Bacteroides	0.000	0.000	0.079	0.057	0.101
	OTUBacteroides_nordii:Bacteroidetes;Bacteroides	0.000	0.001	-0.402	-0.611	-0.194		OTUBacteroides_stercoris:Bacteroidetes;Bacteroides	0.000	0.000	0.431	0.247	0.615
	OTUBacteroides_eggerthii:Bacteroidetes;Bacteroides	0.003	0.013	-0.329	-0.545	-0.114							
	OTUBacteroides_acidifaciens:Bacteroides_xylanisolvens:Bacteroidetes;Bacteroides	0.003	0.015	-0.176	-0.293	-0.059							
	OTUBacteroides_vulgatus:Bacteroidetes;Bacteroides	0.006	0.026	0.172	0.050	0.295							
	OTUBacteroides_acidifaciens:Bacteroidetes;Bacteroides	0.006	0.027	-0.103	-0.177	-0.030							
	Zeller et al. (Stool)							**Zeller et al. (Stool)**					
	Bacteroides sp. 2_1_16	0.003	0.081	0.106	0.035	0.176		OTUBacteroides_coprocola:Bacteroidetes;Bacteroides	0.000	0.000	0.100	0.065	0.134
								OTUBacteroides_vulgatus:Bacteroidetes;Bacteroides	0.014	0.125	-0.112	-0.200	-0.023
								OTUBacteroides_oleiciplenus:Bacteroides_stercorirosoris:Bacteroidetes;Bacteroides	0.021	0.166	-0.245	-0.452	-0.037
								OTUotu2597:Bacteroides_eggerthii:Bacteroides_helcogenes:Bacteroidetes;Bacteroides	0.007	0.073	-0.507	-0.872	-0.142
	Feng et al. (Stool)							**Feng et al. (Stool)**					
	Bacteroides sp. 2_1_16	0.001	0.005	-0.184	-0.289	-0.078		Bacteroides sp. 2_1_16	0.046	0.156	0.113	0.002	0.223
	Bacteroides caccae	0.002	0.013	-0.173	-0.285	-0.062							
	Bacteroides thetaiotaomicron	0.003	0.013	-0.155	-0.256	-0.054							
	Bacteroides sp. 2_1_22	0.005	0.022	-0.127	-0.215	-0.039							
	Bacteroides vulgatus	0.013	0.051	-0.135	-0.242	-0.028							
	Vogtmannn et al. (Stool)							**Vogtmannn et al. (Stool)**					
	Bacteroides sp. 2_1_16	0.000	0.004	0.199	0.102	0.296		Bacteroides phage B40-8	0.000	0.000	0.107	0.101	0.114
	Bacteroides intestinalis	0.003	0.098	-0.089	-0.148	-0.030		Bacteroides caccae	0.012	0.106	-0.117	-0.208	-0.026
	Bacteroides fragilis	0.008	0.145	0.087	0.022	0.152		Bacteroides fragilis	0.049	0.328	-0.085	-0.169	0.000
Coprococcus	**Baxter et al. (Stool)**						**Coprococcus**	**Baxter et al. (Stool)**					
	OTUotu2691:Coprococcus_eutactus:Firmicutes;Coprococcus	0.000	0.000	-0.091	-0.101	-0.081		NONE					
	Zackular et al. (Stool)							**Zackular et al. (Stool)**					
	NONE							NONE					
	Zeller et al. (Stool)							**Zeller et al. (Stool)**					
	OTUotu6965:Coprococcus_eutactus:Firmicutes;Coprococcus	0.000	0.000	-0.084	-0.097	-0.071		NONE					
	Feng et al. (Stool)							**Feng et al. (Stool)**					
	NONE							NONE					
	Vogtmannn et al. (Stool)							**Vogtmannn et al. (Stool)**					
	NONE							NONE					

Supplemental Table 2: Mediation analysis among taxa associatied with obesity and CRC.

Study	Taxa	Change in Prob.[a]	Est	p-value	LL (2.5)	UL (97.5)
baxter	Bacteroides_fluxus	-0.00865	0.00562	0.14969	-0.00203	0.01328
baxter	Howardella_ureilytica	-0.00741	0.00067	0.60715	-0.00188	0.00321
baxter	otu1526:Enterorhabdus_caecimuris:Enterorhabdus_mucosicola	-0.00722	-1.00E-04	0.85142	-0.00117	0.00097
baxter	otu2405:Allobaculum_stercoricanis	-0.00726	9.00E-05	0.87294	-0.00106	0.00125
baxter	otu2695:Eubacterium_coprostanoligenes	-0.00865	0.00562	0.30655	-0.00515	0.01639
baxter	otu2753:Vallitalea_guaymasensis	-0.00785	0.00243	0.64083	-0.00779	0.01266
baxter	otu847:Clostridium_aerotolerans:Clostridium_algidixylanolyticum:Clostridium_saccharolyticum:Clostridium_xylanolyticum:Gracilibacter_thermotolerans	-0.00798	0.00296	0.62175	-0.00881	0.01474
baxter	otu911:Intestinimonas_butyriciproducens	-0.008	0.00304	0.66624	-0.01077	0.01685
feng	Haemophilus parainfluenzae	0.00763	-5.00E-05	0.98819	-0.00621	0.00612
feng	Lactobacillus casei group	0.00753	0.00037	0.88939	-0.00482	0.00556
feng	Prevotella denticola	0.00884	-0.00487	0.7282	-0.03235	0.0226
feng	Prevotella ruminicola	0.00878	-0.00464	0.70382	-0.02859	0.0193
zackular	Bacteroides_eggerthii	-0.01117	-0.00027	0.99158	-0.05084	0.0503
zackular	Bacteroides_nordii	-0.01306	0.0073	0.50709	-0.01427	0.02887
zackular	Bacteroides_salyersiae	-0.01058	-0.00262	0.62977	-0.01325	0.00802
zackular	Gemmiger_formicilis	-0.01153	0.00116	0.72148	-0.0052	0.00751
zackular	otu1157:Alistipes_indistinctus	-0.01127	0.00014	0.98768	-0.01788	0.01816
zackular	otu1202:Clostridium_botulinum:Clostridium_sporogenes	-0.01062	-0.00248	0.69992	-0.01512	0.01015
zackular	otu1257:Intestinimonas_butyriciproducens	-0.01126	9.00E-05	0.98154	-0.00782	0.00801
zackular	otu1316:Ruminococcus_flavefaciens	-0.01065	-0.00236	0.74719	-0.01668	0.01197
zackular	otu1989:Eubacterium_coprostanoligenes	-0.01113	-0.00041	0.90967	-0.0075	0.00668
zackular	otu2005:Alistipes_finegoldii:Alistipes_massiliensis	-0.01097	-0.00108	0.75814	-0.00793	0.00578
zackular	otu327:Prevotella_oris	-0.01158	0.00136	0.77451	-0.00794	0.01066
zackular	otu476:Clostridium_saccharogumia	-0.01117	-0.00026	0.96635	-0.01256	0.01203
zackular	otu508:Caloramator_fervidus:Trigonala_elaeagnus	-0.01108	-0.00064	0.94102	-0.01763	0.01635
zackular	Phascolarctobacterium_succinatutens	-0.01117	-0.00025	0.92066	-0.00523	0.00473
zeller.Stool	Clostridium_bolteae:Clostridium_clostridioforme	-0.00063	0.00049	0.87263	-0.00547	0.00644
zeller.Stool	Dialister_invisus	-0.00037	-0.00052	0.94127	-0.01446	0.01341
zeller.Stool	Dialister_succinatiphilus	-0.00081	0.00121	0.76375	-0.0067	0.00912
zeller.Stool	otu2321:Streptococcus_salivarius:Streptococcus_thermophilus:Streptococcus_vestibularis	0.00046	-0.00388	0.81746	-0.03681	0.02906
zeller.Stool	otu2483:Prevotella_copri	-9E-05	-0.00167	0.80792	-0.01516	0.01182
zeller.Stool	otu4937:Peptococcus_niger	-0.00023	-0.00108	0.76023	-0.00805	0.00588
zeller.Stool	otu834:Prevotella_copri:Prevotella_stercorea	-0.00049	-5.00E-05	0.97294	-0.00321	0.0031
zeller.Stool	Phascolarctobacterium_succinatutens	-0.00054	0.00013	0.99755	-0.08285	0.08311
zeller.Stool	Streptococcus_porcinus:Streptococcus_seminale:Streptococcus_uberis	0.00037	-0.00352	0.62505	-0.01762	0.01059
zeller.Stool	Streptococcus_salivarius	0.00138	-0.00756	0.28432	-0.02139	0.00628
zeller.WGS	Bifidobacterium bifidum	-0.00557	0.00212	0.98433	-0.2097	0.21395
zeller.WGS	Bifidobacterium catenulatum	-0.00443	-0.00244	0.37338	-0.00783	0.00294
zeller.WGS	Streptococcus salivarius	-0.00694	0.00761	0.41392	-0.01065	0.02588
zeller.WGS	Bacteroides sp. 2_1_16	0.05349		1.05494		
zeller.WGS	Streptococcus salivarius	0.05349	0.0257	1.05494	1.02604	-0.0289
zeller.WGS	[Eubacterium] eligens	0.05349		1.05494		
zeller.WGS	Bifidobacterium bifidum	0.05349	0.0312	1.05494	1.03169	-0.02325

[a.] qunatificaiton of the change in OR from model 1 to model 2.

Supplemental Table 3: Mediation Cross Tabulation

Study	Bacteria	CRC_Bact_Present	CRC_Bact_NotPresent	nonCRC_Bact_Present	nonCRC_Bact_NotPresent
baxter	otu2695:Eubacterium_coprostanoligenes	102	54	267	63
baxter	otu911:Intestinimonas_butyriciproducens	209	87	160	30
baxter	otu2405:Allobaculum_stercoricanis	91	32	278	85
baxter	Howardella_ureilytica	70	28	299	89
baxter	otu1526	83	28	286	89
baxter	otu2753:Vallitalea_guaymasensis	181	45	188	72
baxter	otu847	237	60	132	57
baxter	Bacteroides_fluxus	65	3	304	114
zackular	otu508:Caloramator_fervidus	32	13	26	11
zackular	Gemmiger_formicilis	52	20	6	4
zackular	otu1989:Eubacterium_coprostanoligenes	21	9	37	15
zackular	otu2005	9	4	49	20
zackular	Phascolarctobacterium_succinatutens	16	6	42	18
zackular	otu1257:Intestinimonas_butyriciproducens	30	13	28	11
zackular	Bacteroides_salyersiae	6	2	52	22
zackular	otu476:Clostridium_saccharogumia	44	17	14	7
zackular	Bacteroides_eggerthii	25	5	33	19
zackular	otu1157:Alistipes_indistinctus	25	9	33	15
zackular	otu1202	10	4	48	20
zackular	Bacteroides_nordii	24	6	34	18
zackular	otu327:Prevotella_oris	15	8	43	16
zackular	otu1316:Ruminococcus_flavefaciens	13	8	45	16
zeller.Stool	Phascolarctobacterium_succinatutens	42	21	44	20
zeller.Stool	otu4937:Peptococcus_niger	18	11	68	30
zeller.Stool	otu2483:Prevotella_copri	29	13	57	28
zeller.Stool	otu834	15	6	71	35
zeller.Stool	Dialister_invisus	29	10	57	31
zeller.Stool	Streptococcus_salivarius	9	1	77	40
zeller.Stool	Streptococcus_porcinus	6	2	80	39
zeller.Stool	Dialister_succinatiphilus	20	8	66	33
zeller.Stool	Clostridium_bolteae	53	26	33	15
zeller.Stool	otu2321	58	21	28	20
feng	Ruminococcus sp. 5_1_39BFAA	110	46		
feng	Prevotella denticola	73	40	37	6
feng	Lactobacillus casei group	81	36	29	10
feng	Haemophilus parainfluenzae	84	36	26	10
feng	Prevotella ruminicola	72	39	38	7
feng	Coprobacillus sp. D7	110	46		
zeller.WGS	Bifidobacterium catenulatum	103	88	2	1
zeller.WGS	Bacteroides sp. 2_1_16	105	89		
zeller.WGS	Streptococcus salivarius	105	88		
zeller.WGS	[Eubacterium] eligens	105	89		
zeller.WGS	Bifidobacterium bifidum	105	88		

ium_aerotolerans:Clostridium_algidixylanolyticum:Clostridium_saccharolyticum:Clostridium_
s_thermophilus:Streptococcus_vestibularis otu508:Caloramator_fervidus:Trigonala_elaeagnus Clostridium_bolteae:Clostridium_clostridioforme Streptococcus_porcinus:Streptococcus_seminale:Streptococcus_uberis
lla_copri:Prevotella_stercorea

Supplemental Table 4. Taxa with Q values <0.2 after performing mediation analysis between PooledDS and PooledBMI

Baxter	Zackular	Zeller (16S)	Feng	Zeller (WGS)	Vogtmann
OTUotu2695	OTUotu508	OTUPhascolarctobacterium_succinatutens	Ruminococcus sp. 5_1_39BFAA	Bifidobacterium catenulatum	None
OTUotu911	OTUGemmiger_formicilis	OTUotu4937	Prevotella denticola	Bacteroides sp. 2_1_16	
OTUotu2405	OTUotu1989	OTUotu2483	Lactobacillus casei group	Streptococcus salivarius	
OTUHowardella_ureilytica	OTUotu2005	OTUotu834	Haemophilus parainfluenzae	[Eubacterium] eligens	
OTUotu1526	OTUPhascolarctobacterium_succinatutens	OTUDialister_invisus	Prevotella ruminicola	Bifidobacterium bifidum	
OTUotu2753	OTUotu1257	OTUStreptococcus_salivarius	Coprobacillus sp. D7		
OTUotu847	OTUBacteroides_salyersiae	OTUStreptococcus_porcinus			
OTUBacteroides_fluxus	OTUotu476	OTUDialister_succinatiphilus			
	OTUBacteroides_eggerthii	OTUClostridium_bolteae			
	OTUotu1157	OTUotu2321			
	OTUotu1202				
	OTUBacteroides_nordii				
	OTUotu327				
	OTUotu1316				