

SUPPLEMENTARY INFORMATION

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 White², R. Noah Padgett³, Brittany G Perrotta⁴, Gregory D Jenkins⁵, Nicholas Chia^{5,6}, Jun Chen⁶.

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.

Corresponding Author: K. Leigh Greathouse, One Bear Place 97346, Waco, TX, 76798-7346, Email: Leigh_Greathouse@baylor.edu

Contents

Supporting Methods

Sample Population.....	page 3
Processing of Microbial Reads and Calculation of Diversity.....	page 3
Prediction of Metagenomic Pathways.....	page 4
Statistical Analyses.....	page 5
References.....	page 7

Supplementary Figures

Figure S1: Probability of having CRC using alpha diversity as a predictor among individuals with obesity.....	page 9
Figure S2: Alpha diversity by BMI in individuals with or without CRC.....	page 10
Figure S3: Beta diversity in individuals with or without obesity and with or without CRC..	page 11
Figure S4: Differential abundance of taxa associated with obesity and CRC.....	page 12
Figure S5: Ability of CRC-associated taxa to predict CRC among individuals with obesity...	page 13
Figure S6: PRISMA Search Diagram.....	page 14

Supplementary Tables

Table 1: Summary of species level differential abundance comparing obese or non-obese controls to obese or non-obese CRC cases.....	page 15
Table 2: Identification of taxa that mediate the association between obesity and CRC.....	page 19
Table 3: Overall presence or absence of each taxa in CRC cases and controls that are associated with obesity and CRC.....	page 20
Table 4: Taxa with Q values <0.2 after performing mediation analysis between PooledDS and PooledBMI samples.....	page 21

Supplementary Methods:

Sample Population

We used the PRISMA guidelines and applied the following search terms in PubMed to initially identify studies for analysis (((("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 (Figure S6).

Processing of Microbial Reads and Calculation of Diversity

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 v1.9.0. 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 (1-3)

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 (4, 5).

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

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. Precisely, we are estimating the following model:

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

Where the regression coefficients are modeled by study characteristics. For example, $\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Sequencing Method}) + \gamma_{02}(\text{Variable Region}) + u_{0j}$; which defines how the model parameters 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 (7). Due to the small 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 (8).

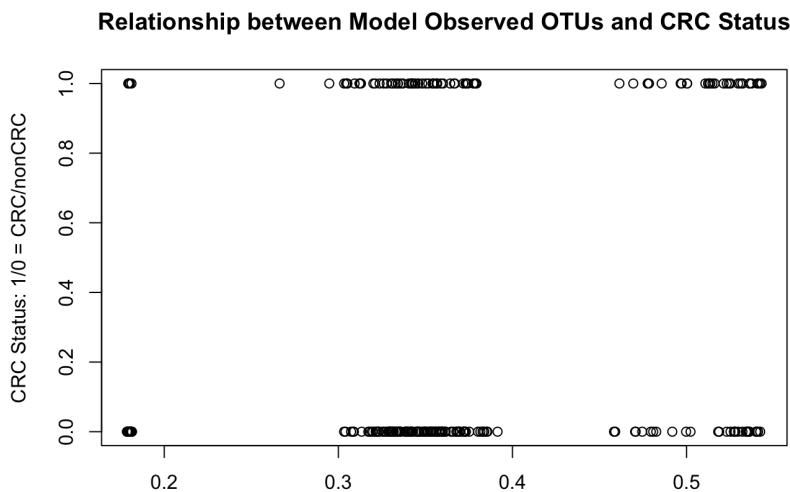
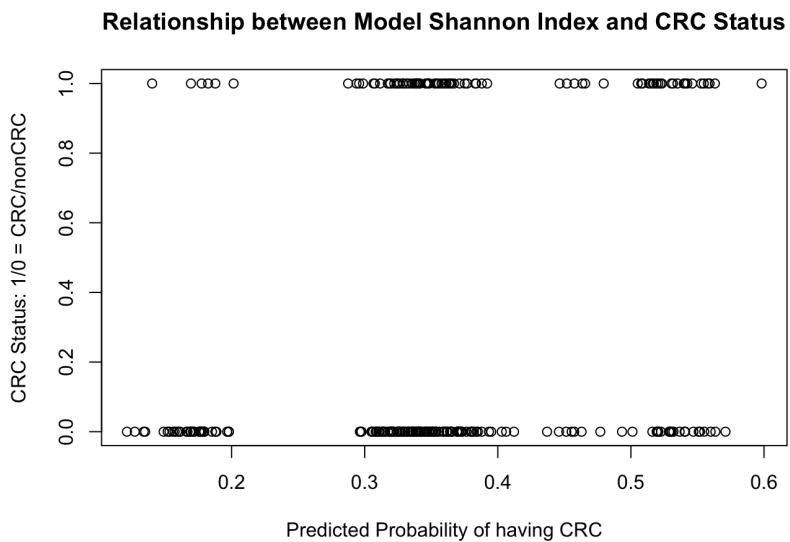
Mediation analyses were also conducted. The goal of these analyses is to uncover if the relationship between BMI and CRC status is mediated by bacteria present. First, bacteria were dummy coded for presence or not for all samples (dummy coded solely for whether an individual has a given bacteria or not, 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 (9). This model is estimated for the presence of each bacterium. Lastly, the change in the odds 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. Last, we derived an overall mediation affect using the q-values from Pooled BMI (association between BMI and microbiome using all samples only, adjusting for disease status, sex and age) and Pooled DS data sets (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.

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.

References:

1. Drewes JL, White JR, Dejea CM, et al. High-resolution bacterial 16S rRNA gene profile meta-analysis and biofilm status reveal common colorectal cancer consortia. *NPJ Biofilms Microbiomes* 2017;3:34.
2. Daugman N, Seekatz AM, Greathouse KL, et al. High-resolution profiling of the gut microbiome reveals the extent of Clostridium difficile burden. *NPJ Biofilms Microbiomes* 2017;3:35.
3. Guerrero-Preston R, White JR, Godoy-Vitorino F, et al. 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* 2017;8:110931-110948.
4. Byrd AL, Perez-Rogers JF, Manimaran S, et al. Clinical PathoScope: rapid alignment and filtration for accurate pathogen identification in clinical samples using unassembled sequencing data. *BMC Bioinformatics* 2014;15:262.
5. Oh J, Byrd AL, Deming C, et al. Biogeography and individuality shape function in the human skin metagenome. *Nature* 2014;514:59-64.
6. Dejea CM, Fathi P, Craig JM, et al. Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science* 2018;359:592-597.
7. Bates D, Mächler M, Bolker B, et al. Fitting Linear Mixed-Effects Models Using lme4. *2015* 2015;67:48.
8. meta: An R package for meta-analysis [R News: R News; 2007].
9. Rosseel Y. lavaan: An R Package for Structural Equation Modeling. *2012* 2012;48:36.

A**B****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.

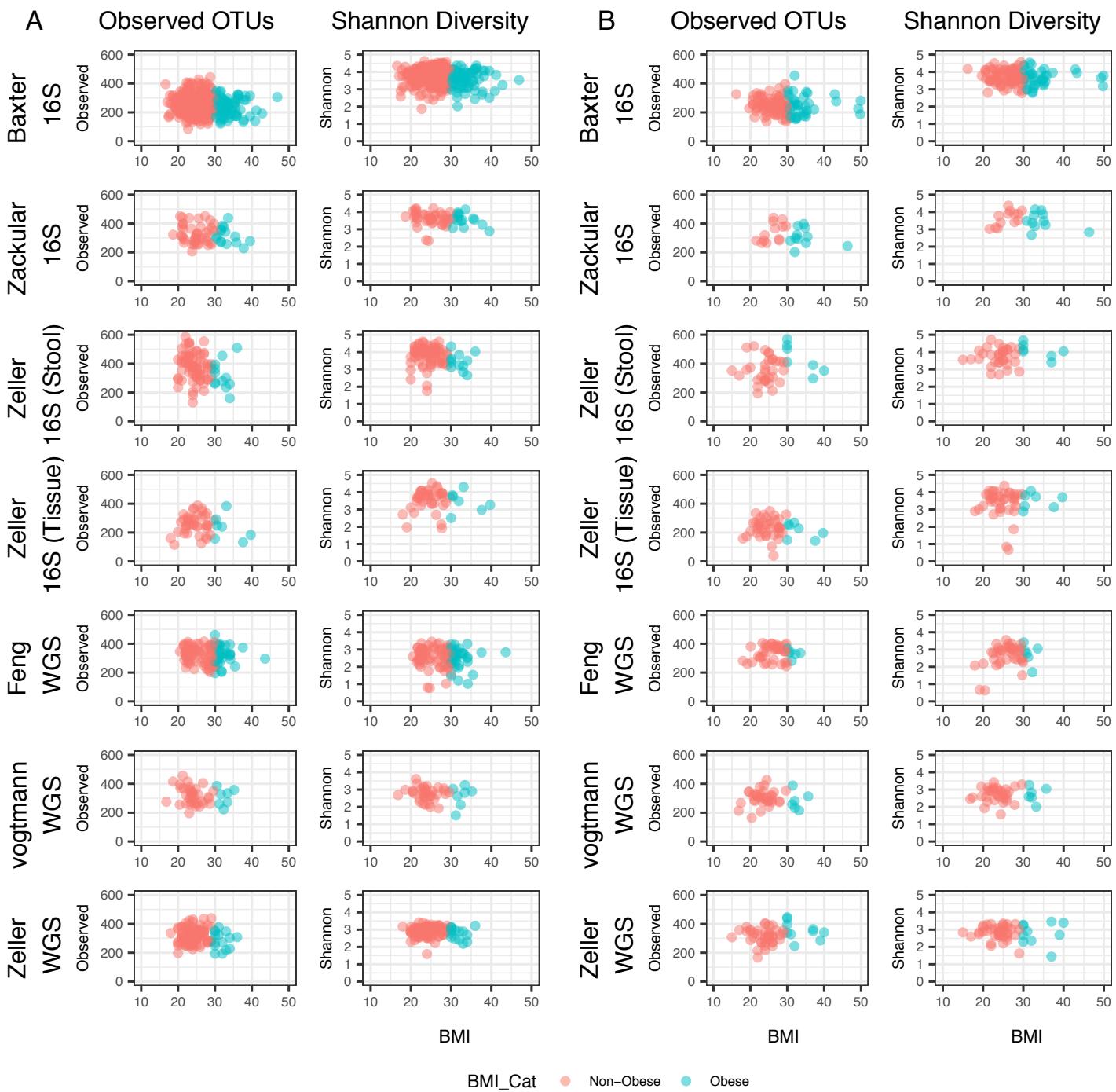


Figure S2: Alpha diversity by BMI in individuals with or without CRC.

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.

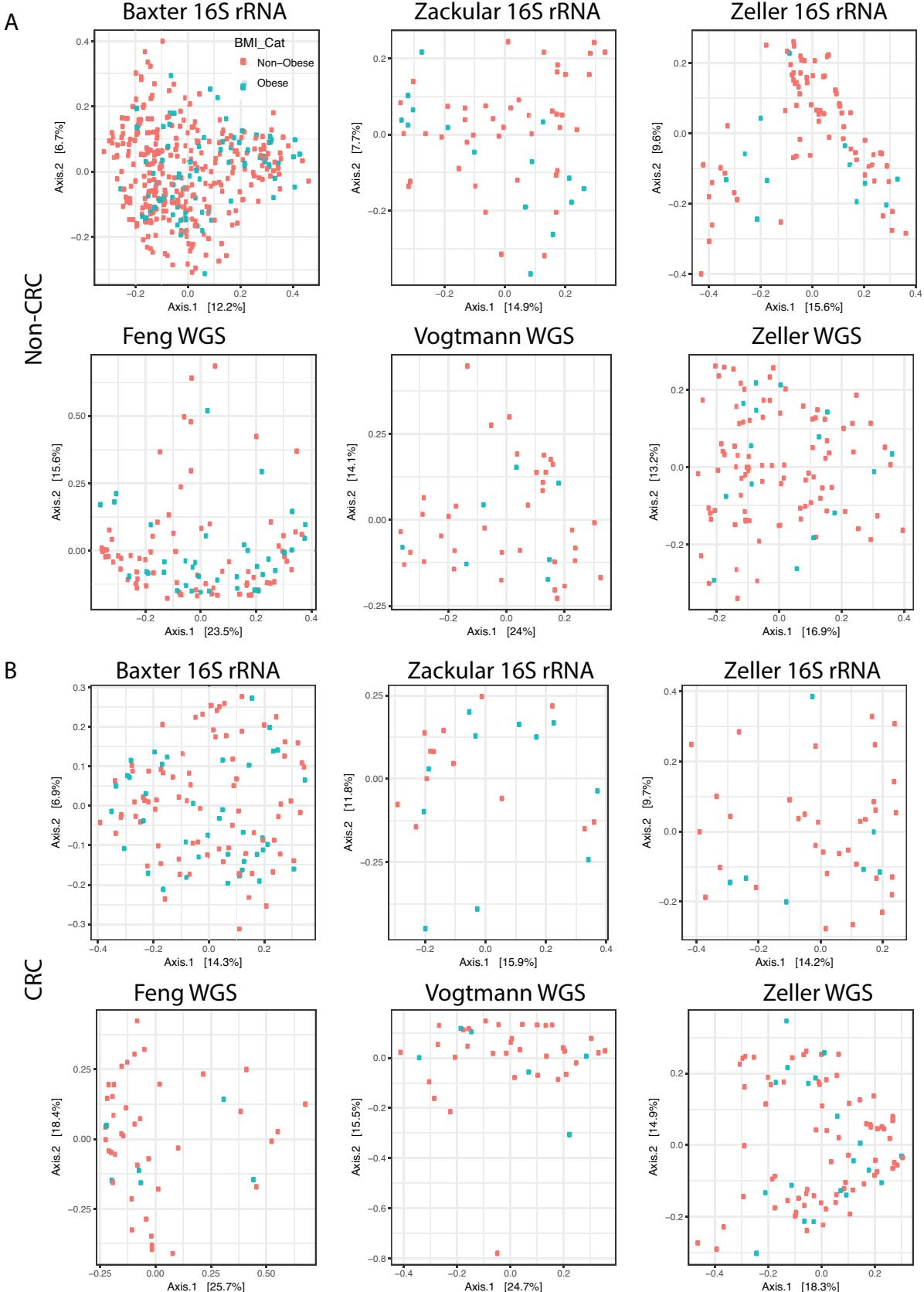


Figure S3: Beta diversity in individuals with or without obesity and with or without CRC.

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.

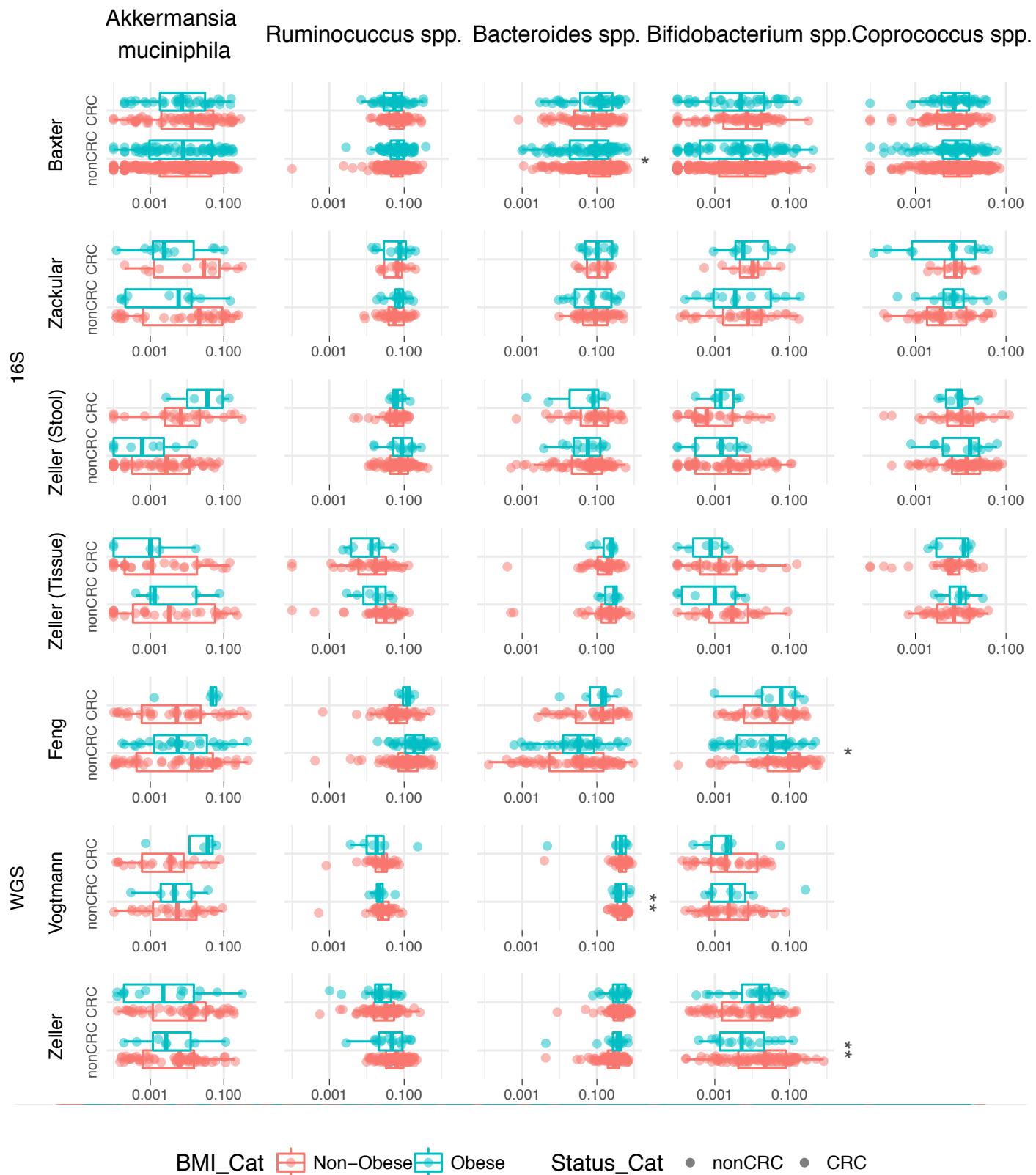


Figure S4: Differential abundance of taxa associated with obesity and CRC taxa.

OTUs (16S rRNA) or species (WGS) log 10 scale relative abundance of *Ruminococcus* 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).

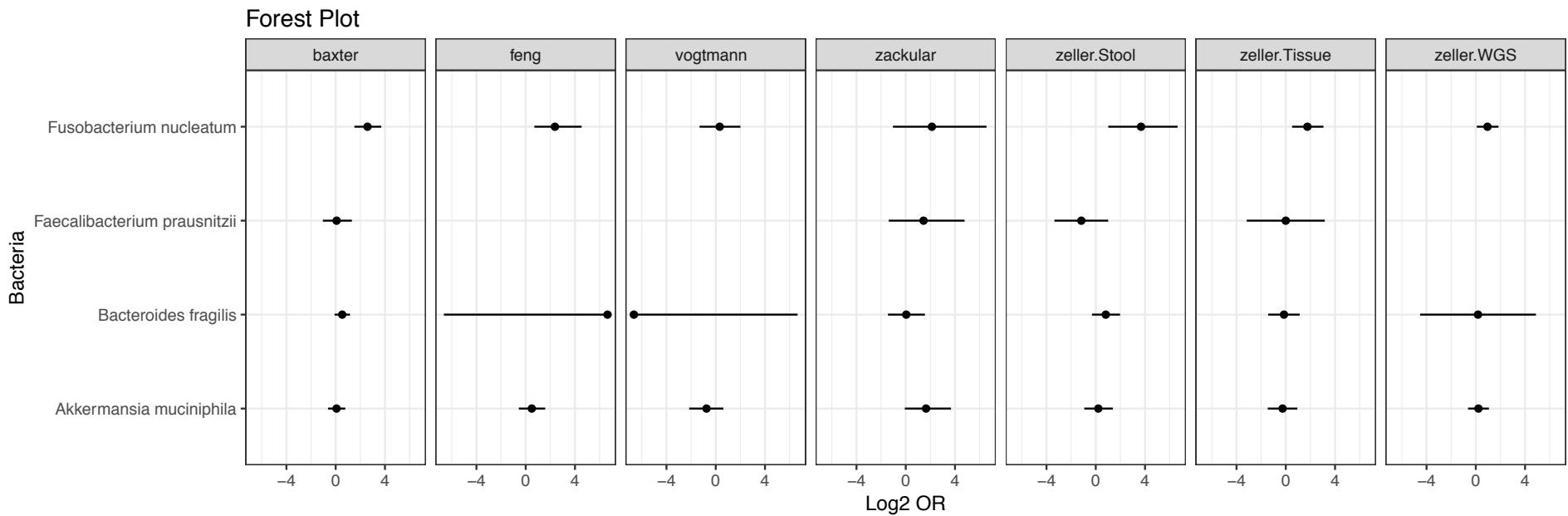


Figure S5: Ability of CRC-associated taxa to predict CRC among individuals with obesity.

For each species identified from previous CRC microbiome studies, *F. nucleatum*, *F. prausnitzii*, *B. fragilis*, or *A. muciniphila*, the log₂ odd ratio was calculated for individuals with obesity to determine odds of being classified as having CRC.

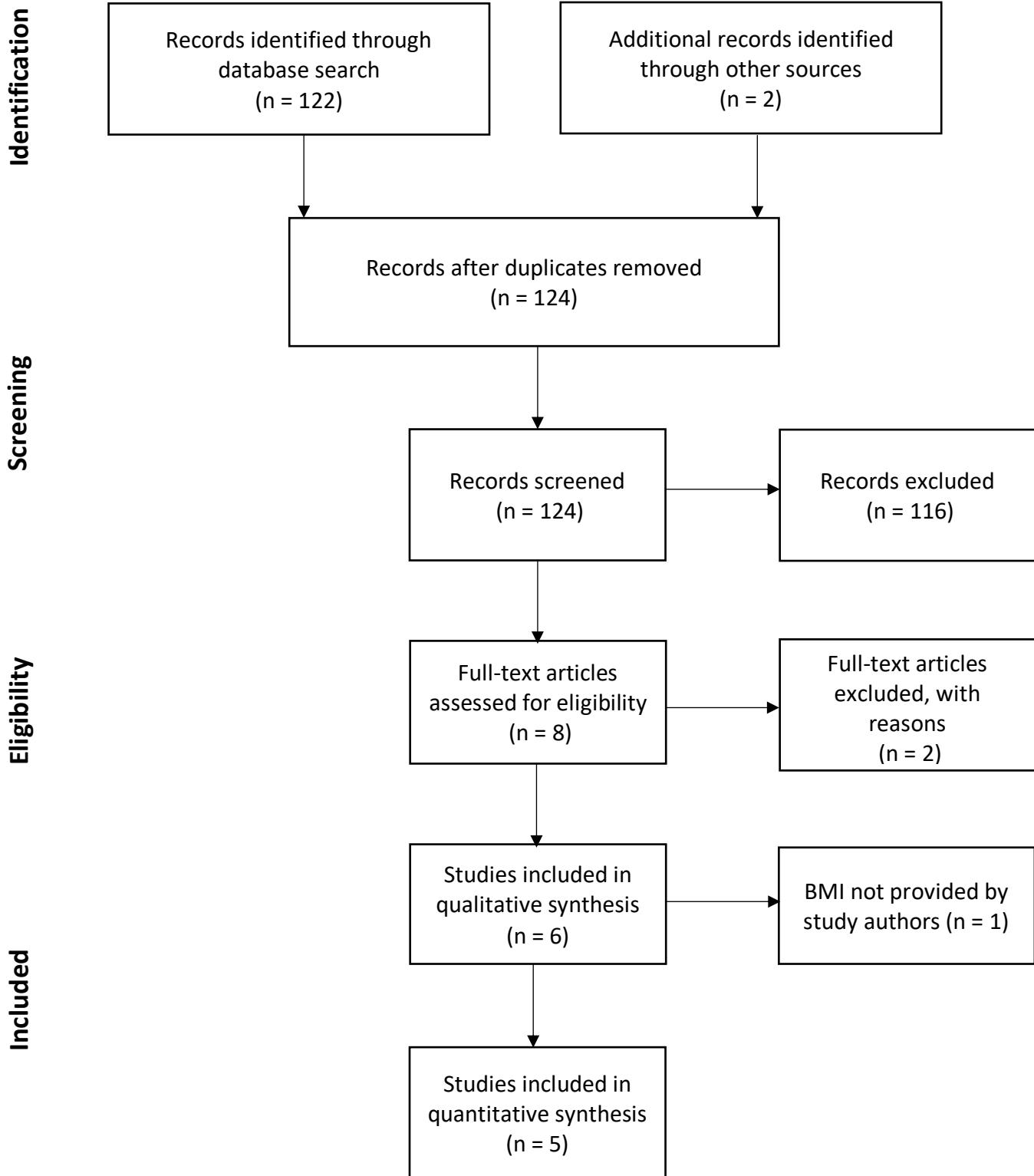


Figure S6:

PRISMA Search Diagram. Flow chart of selection criteria for identification, screening, eligibility and final studies included.

Supplemental Table 1: Summary of species level differential abundance comparing obese or non-obese controls to obese or non-obese CRC cases

Genus	Reference (Sample)	Non-CRC Obese v Non-Obese				
		Pvalue	Qvalue	BMI	2.50%	97.50%
Akkermansia	<i>Species</i>					
	Baxter et al. (Stool)					
	NONE					
	Zackular et al. (Stool)					
	OTU Akkermansia_muciniphila:Verrucomicrobia;Akkermansia	0.001	0.004	-0.240	-0.378	-0.101
Ruminococcus	Zeller et al. (Stool)					
	NONE					
	Feng et al. (Stool)					
	NONE					
	Vogtmann et al. (Stool)					
Ruminococcus	NONE					
	Baxter et al. (Stool)					
	OTU Ruminococcus_lactaris:Firmicutes;Ruminococcus	0.001	0.019	0.060	0.024	0.096
	Zackular et al. (Stool)					
	OTU otu1316:Ruminococcus_flavefaciens:Firmicutes;Ruminococcus	0.000	0.000	-0.151	-0.171	-0.131
	OTU otu1473:Clostridium_methylpentosum:Ruminococcus_albus:Firmicutes;unassigned	0.000	0.002	-0.453	-0.700	-0.205
	OTU otu148:Ruminococcus_bromii:Firmicutes;Ruminococcus	0.008	0.033	-0.287	-0.499	-0.075
	Zeller et al. (Stool)					
	OTU Ruminococcus_champanellensis:Firmicutes;Ruminococcus	0.000	0.007	-0.234	-0.358	-0.109
	OTU otu1488:Ruminococcus_bromii:Firmicutes;Ruminococcus	0.003	0.042	-0.151	-0.249	-0.052
	OTU otu4879:Blautia_faecis:Clostridium_boliensis:Clostridium_celerecrescens:Clostridium_hatthewayi:Clostridium_saccharolyticum:Clostridium_sphenoides:Ruminococcus_lactaris:Firmicutes;unassigned	0.003	0.043	-0.227	-0.377	
	OTU otu1:Ruminococcus_bromii:Firmicutes;Ruminococcus	0.004	0.051	-0.419	-0.701	-0.137
	OTU otu6413: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)					
	NONE					
	Vogtmann et al. (Stool)					
	Ruminococcus sp. 5_1_39BFAA	0.860	0.983	0.004	-0.042	0.050
Bifidobacterium	Baxter et al. (Stool)					
	NONE					
	Zackular et al. (Stool)					
	NONE					
	OTU Bifidobacterium_catenuatum:Bifidobacterium_kashiwanohense:Bifidobacterium_pseudocatenuatum:Actinobacteria;Bifidobacterium	0.656	0.769	0.054	-0.184	0.292
	Zeller et al. (Stool)					
	OTU Bifidobacterium_catenuatum:Bifidobacterium_kashiwanohense:Bifidobacterium_pseudocatenuatum:Actinobacteria;Bifidobacterium	0.008	0.082	-0.272	-0.472	-0.072
	Bifidobacterium_catenuatum	0.005	0.088	-0.166	-0.281	-0.051
	Bifidobacterium_adolescentis	0.028	0.192	-0.140	-0.265	-0.015
	Feng et al. (Stool)					
	Bifidobacterium_dentium	0.021	0.067	0.144	0.022	0.266
	Vogtmann et al. (Stool)					
	Bifidobacterium_adolescentis	0.877	0.985	0.014	-0.163	0.191
	Bifidobacterium_bifidum	0.218	0.581	0.068	-0.040	0.177
	Bifidobacterium_longum	0.200	0.581	0.075	-0.040	0.190
	Bifidobacterium_catenuatum	0.009	0.145	0.211	0.052	0.370

P-values - negative binomial regression; BMI used as a continuous variable controlling for age and sex.

Q-values - Benjamini-Hochberg-based false discovery rate (significant at <0.2)

Standardized mean differences among studies calculated using Hedge's g

Supplemental Table 1: Summary of species level differential abundance comparing obese or non-obese controls to obese or non-obese CRC cases (cont.)

Genus	Reference (Sample)	Non-CRC Obese v Non-Obese				
		Pvalue	Qvalue	BMI	2.50%	97.50%
Bacteroides	Species					
	Baxter et al. (Stool)					
	OTUBacteroides_eggerthii:Bacteroidetes;Bacteroides	0.009	0.099	-0.171	-0.300	-0.042
	OTUBacteroides_fluxus:Bacteroidetes;Bacteroides	0.000	0.000	0.049	0.036	0.063
	Zackular et al. (Stool)					
	OTUBacteroides_salyersiae:Bacteroidetes;Bacteroides	0.000	0.000	0.147	0.114	0.180
	OTUBacteroides_nordii:Bacteroidetes;Bacteroides	0.000	0.001	-0.402	-0.611	-0.194
	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)					
	Bacteroides sp. 2_1_16	0.003	0.081	0.106	0.035	0.176
	Feng et al. (Stool)					
	Bacteroides sp. 2_1_16	0.001	0.005	-0.184	-0.289	-0.078
	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
	Vogtmann et al. (Stool)					
	Bacteroides sp. 2_1_16	0.000	0.004	0.199	0.102	0.296
	Bacteroides intestinalis	0.003	0.098	-0.089	-0.148	-0.030
	Bacteroides fragilis	0.008	0.145	0.087	0.022	0.152
Coprococcus	Baxter et al. (Stool)					
	OTUotu2691:Coprococcus_eutactus:Firmicutes;Coprococcus	0.000	0.000	-0.091	-0.101	-0.081
	Zackular et al. (Stool)					
	NONE					
	Zeller et al. (Stool)					
	OTUotu6965:Coprococcus_eutactus:Firmicutes;Coprococcus	0.000	0.000	-0.084	-0.097	-0.071
	Feng et al. (Stool)					
	NONE					
	Vogtmann et al. (Stool)					
	NONE					

P-values - negative binomial regression; BMI used as a continuous variable controlling for age and sex.

Q-values - Benjamini-Hochberg-based false discovery rate (sig. set at <0.2)

Standardized mean differences among studies calculated using Hedge's g

Supplemental Table 1: Summary of species level differential abundance comparing obese or non-obese controls to obese or non-obese CRC cases (cont.)

Genus	Reference (Sample)	CRC - Ob v Non-Ob				
		Pvalue	Qvalue	BMI	2.50%	97.50%
Akkermansia	<i>Species</i>					
	Baxter et al. (Stool)					
	OTU Akkermansia_muciniphila:Verrucomicrobia;Akkermansia	0.008	0.090	-0.069	-0.120	-0.018
	Zackular et al. (Stool)					
	NONE					
	Zeller et al. (Stool)					
	OTU otu3490:Akkermansia_muciniphila:Verrucomicrobia;Akkermansia	0.000	0.004	-0.104	-0.159	-0.050
	Feng et al. (Stool)					
	Akkermansia muciniphila	0.024	0.115	0.183	0.024	0.342
Ruminococcus	Vogtmann et al. (Stool)					
	Baxter et al. (Stool)					
	NONE					
	Zackular et al. (Stool)					
	OTU Ruminococcus_callidus:Firmicutes;Ruminococcus	0.000	0.000	-0.110	-0.129	-0.091
	OTU otu1072:Ruminococcus_callidus:Ruminococcus_champanellensis:	0.000	0.000	-0.071	-0.098	-0.044
	Ruminococcus_flavefaciens:Firmicutes;Ruminococcus					
	OTU Ruminococcus_champanellensis:Firmicutes;Ruminococcus	0.002	0.010	-0.233	-0.381	-0.085
	OTU otu1794:Ruminococcus_flavefaciens:Firmicutes;Ruminococcus	0.013	0.050	-0.807	-1.446	-0.169
	OTU otu1473:Clostridium_methylpentosum:Ruminococcus_albus:Firm	0.000	0.002	0.337	0.156	0.519
	icutes;unassigned					
	Zeller et al. (Stool)					
	OTU Ruminococcus_callidus:Firmicutes;Ruminococcus	0.028	0.199	0.175	0.019	0.331
	Feng et al. (Stool)					
	NONE					
	Vogtmann et al. (Stool)					
	[Ruminococcus] gnatus	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)					
	NONE					
	Zackular et al. (Stool)					
	OTU Bifidobacterium_adolescentis:Bifidobacterium_stercoris:Actinoba	0.000	0.000	0.025	0.019	0.032
	OTU Bifidobacterium_bifidum:Actinobacteria;Bifidobacterium	0.000	0.000	0.048	0.030	0.066
	OTU Bifidobacterium_catenulatum:Bifidobacterium_kashiwanohense:	0.000	0.002	-0.033	-0.051	-0.015
	Bifidobacterium_pseudocatenulatum:Actinobacteria;Bifidobacterium					
	Zeller et al. (Stool)					
	Bifidobacterium catenulatum	0.008	0.096	-0.119	-0.207	-0.032
	Bifidobacterium pseudocatenulatum	0.012	0.129	-0.095	-0.170	-0.021
	Feng et al. (Stool)					
	Bifidobacterium catenulatum	0.001	0.005	-0.298	-0.466	-0.130
	Bifidobacterium longum	0.037	0.135	0.112	0.007	0.217
	Vogtmann et al. (Stool)					
	Bifidobacterium adolescentis	0.305	0.675	0.100	-0.091	0.291
	Bifidobacterium longum	0.311	0.675	-0.056	-0.163	0.052

P-values - negative binomial regression; BMI used as a continuous variable controlling for age and sex.

Q-values - Benjamini-Hochberg-based false discovery rate (significant at <0.2)

Standardized mean differences among studies calculated using Hedge's g

Supplemental Table 1: Summary of species level differential abundance comparing obese or non-obese controls to obese or non-obese CRC cases (cont.)

Genus	Reference (Sample)	CRC - Ob v Non-Ob				
		Pvalue	Qvalue	BMI	2.50%	97.50%
Bacteroides	Species					
	Baxter et al. (Stool)					
	OTUBacteroides_plebeius:Bacteroidetes;Bacteroides	0.000	0.000	0.037	0.023	0.051
	OTUBacteroides_finegoldii:Bacteroidetes;Bacteroides	0.002	0.045	-0.164	-0.267	-0.061
	Zackular et al. (Stool)					
	OTUBacteroides_nordii:Bacteroidetes;Bacteroides	0.000	0.000	0.079	0.057	0.101
	OTUBacteroides_stercoris:Bacteroidetes;Bacteroides	0.000	0.000	0.431	0.247	0.615
	Zeller et al. (Stool)					
	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_stercorisoris: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)					
	Bacteroides sp. 2_1_16	0.046	0.156	0.113	0.002	0.223
	Vogtmann et al. (Stool)					
	Bacteroides phage B40-8	0.000	0.000	0.107	0.101	0.114
	Bacteroides caccae	0.012	0.106	-0.117	-0.208	-0.026
	Bacteroides fragilis	0.049	0.328	-0.085	-0.169	0.000
Coprococcus	Baxter et al. (Stool)					
	NONE					
	Zackular et al. (Stool)					
	NONE					
	Zeller et al. (Stool)					
	NONE					
	Feng et al. (Stool)					
	NONE					
	Vogtmann et al. (Stool)					
	NONE					

P-values - negative binomial regression; BMI used as a continuous variable controlling for age and sex.

Q-values - Benjamini-Hochberg-based false discovery rate (sig. set at <0.2)

Standardized mean differences among studies calculated using Hedge's g

Supplemental Table 2: Identification of taxa that mediate the association between 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		-0.00722				
	otu1526:Enterorhabdus_caecimuris:Enterorhabdus_mucosicola		-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_algidiyylanolyticum: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
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_catenuatum	-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 quantification of the change in OR from model 1 to model 2.

Supplemental Table 3: Overall presence or absence of each taxa in CRC cases and controls that are associated with obesity and CRC.

Study	Bacteria	CRC_Bact_Present	CRC_Bact_NotPresent	nonCRC_Bact_Present	nonCRC_Bact_NotPresent	Total NonCRC Controls	Total CRC Cases
Baxter	otu2695:Eubacterium_coprostanoligenes	102	54	267	63	330	156
Baxter	otu911:Intestinimonas_butyriciproducens	209	87	160	30	190	296
Baxter	otu2405:Allobaculum_stercoricanis	91	32	278	85	363	123
Baxter	Howardella_ureilytica	70	28	299	89	388	98
Baxter	otu1526	83	28	286	89	375	111
Baxter	otu2753:Vallitalea_guaymasensis	181	45	188	72	260	226
Baxter	otu847	237	60	132	57	189	297
Baxter	Bacteroides_fluxus	65	3	304	114	418	68
Zackular	otu508:Caloramator_fervidus	32	13	26	11	37	45
Zackular	Gemmiger_formicilis	52	20	6	4	10	72
Zackular	otu1989:Eubacterium_coprostanoligenes	21	9	37	15	52	30
Zackular	otu2005	9	4	49	20	69	13
Zackular	Phascolarctobacterium_succinatutens	16	6	42	18	60	22
Zackular	otu1257:Intestinimonas_butyriciproducens	30	13	28	11	39	43
Zackular	Bacteroides_salyersiae	6	2	52	22	74	8
Zackular	otu476:Clostridium_saccharogumia	44	17	14	7	21	61
Zackular	Bacteroides_eggerthii	25	5	33	19	52	30
Zackular	otu1157:Alistipes_indistinctus	25	9	33	15	48	34
Zackular	otu1202	10	4	48	20	68	14
Zackular	Bacteroides_nordii	24	6	34	18	52	30
Zackular	otu327:Prevotella_oris	15	8	43	16	59	23
Zackular	otu1316:Ruminococcus_flavifaciens	13	8	45	16	61	21
Zeller.Stool	Phascolarctobacterium_succinatutens	42	21	44	20	64	63
Zeller.Stool	otu4937:Peptococcus_niger	18	11	68	30	98	29
Zeller.Stool	otu2483:Prevotella_copri	29	13	57	28	85	42
Zeller.Stool	otu834	15	6	71	35	106	21
Zeller.Stool	Dialister_invisus	29	10	57	31	88	39
Zeller.Stool	Streptococcus_salivarius	9	1	77	40	117	10
Zeller.Stool	Streptococcus_porcinus	6	2	80	39	119	8
Zeller.Stool	Dialister_succinatiphilus	20	8	66	33	99	28
Zeller.Stool	Clostridium_bolteae	53	26	33	15	48	79
Zeller.Stool	otu2321	58	21	28	20	48	79
Feng	Ruminococcus sp. 5_1_39BFAA	110	46	0	0	0	156
Feng	Prevotella_denticola	73	40	37	6	43	113
Feng	Lactobacillus_casei group	81	36	29	10	39	117
Feng	Haemophilus_parainfluenzae	84	36	26	10	36	120
Feng	Prevotella_ruminicola	72	39	38	7	45	111
Feng	Coprobacillus sp. D7	110	46	0	0	0	156
Zeller.WGS	Bifidobacterium_catenulatum	103	88	2	1	3	191
Zeller.WGS	Bacteroides sp. 2_1_16	105	89	0	0	0	194
Zeller.WGS	Streptococcus_salivarius	105	88	0	0	0	193
Zeller.WGS	[Eubacterium]_eligens	105	89	0	0	0	194
Zeller.WGS	Bifidobacterium_bifidum	105	88	0	0	0	193

otu1526:Enterorhabdus_caecimuris:Enterorhabdus_mucosicola otu847:Clostridium_aerotolerans:Clostridium_xylanolyticum:Clostridium_saccharolyticum:Clostridium_xylanolyticum:Gracilibacter_thermotolerans otu2005:Alistipes_finegoldii:Alistipes_massiliensis otu1202:Clostridium_botulinum:Clostridium_sporogenes
otu834:Prevotella_copri:Prevotella_stercorea

Supplemental Table 4. Taxa with Q values <0.2 after performing mediation analysis between all CRC and normal samples (PooledDS) and all samples (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					

Note: Pooled BMI (association between BMI and microbiome using all samples only, adjusting for disease status, sex, and age); Pooled DS (association of disease status with microbiome using CRC and normal samples, adjusting for BMI, sex, and age)