

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

K Leigh Greathouse,^{1,2} James Robert White,³ R Noah Padgett,⁴ Brittany G Perrotta,² Gregory D Jenkins,^{5,6} Nicholas Chia,^{5,6,7} Jun Chen⁷

To cite: Greathouse KL, White JR, Padgett RN, *et al*. Gut microbiome meta-analysis reveals dysbiosis is independent of body mass index in predicting risk of obesity-associated CRC. *BMJ Open Gastro* 2019;**6**:e000247. doi:10.1136/bmjgast-2018-000247

Received 10 September 2018
Revised 13 December 2018
Accepted 27 December 2018

ABSTRACT

Objective 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 in 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 individuals who are not obese.

Design 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, which allowed us to achieve species-level information to compare with WGS.

Results Characterisation 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 diversity 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.

Conclusion Overall, microbial dysbiosis was not a significant factor in explaining the higher risk of colon cancer among individuals with obesity.

INTRODUCTION

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 USA (SEER). Several epidemiological studies demonstrate that adult obesity increases the risk of colon cancer 1.2-fold to 2-fold, with obesity accounting for 14%–35% of total colon cancer incidence.^{1 3–5} 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 aetiology of CRC: obesity, inflammation and the microbiome.

Several studies indicate that specific microbial taxa are playing a role in the aetiology of colon cancer.^{6–9} 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 analysing the microbiome of the gastrointestinal tract (GI). Several studies have found colon cancer-associated microbiota in precancerous colon tissue (adenomas) and have used the microbiome to distinguish precancerous adenomas from CRC, though with variable rates of accuracy.^{10–12} Further, specific bacteria have been identified as promoters in colon cancer development, including enterotoxigenic *Bacteroides fragilis* (ETBF) and *Fusobacterium nucleatum*.^{8 9 13–15}

Multiple lines of evidence demonstrate that both diet and obesity can significantly alter the microbiome.^{16–22} A seminal study illustrating the impact of the microbiome on obesity, transferred the faecal 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 humanised mice.²³ When examining



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr K Leigh Greathouse;
Leigh_Greathouse@baylor.edu



microbiota and subsequent changes in metabolism after faecal 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.^{20 23} More research, however, is necessary to identify the microbiome–host relationship in individuals with obesity.^{24 25}

Chronic inflammation is a hallmark of both obesity and CRC aetiology. Obesity is characterised by proinflammatory adipose tissue macrophages that secrete high levels of IL-17, a cytokine which is also induced by ETBF in murine models of colon cancer.^{9 26 27} Given the reciprocal relationship between the microbiome and the immune system, it is logical to hypothesise 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 proinflammatory factors.¹⁹ In a separate model of colon cancer (K-ras^{G12Dint}), faecal transfer from high-fat fed mice with intestinal tumours to genetically susceptible mice on a standard diet replicated the disease phenotype.²⁸ Thus, it appears that a high fat diet may be sufficient to change the microbiome into a tumour-promoting community independent of obesity and glucose response. 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 used 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 (whole

genome sequencing [WGS]), we processed the 16S rRNA and WGS reads and derived a taxonomic profile from each of the samples. 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. Together, five 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 and

online supplementary material 1). Three of these studies conducted 16S rRNA sequencing on stool or tissue and three conducted WGS on stool or tissue, with one using 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 (V.1.9.0) plus the algorithm Resphera Insight for 16S rRNA sequencing or Pathoscope (V.1.0) 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. Prior to Resphera Insight and statistical analysis we rarefied the 16S rRNA sequencing data to the sample with the lowest number of reads; Zackular 25 000 per sample, Zeller 16S tissue and stool 10 000 per sample, and Baxter 10 000 per sample. Details regarding sequencing methods and variable regions amplified for each data set are listed in table 1 and online supplementary material 1.

Statistical analyses

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 Operational Taxonomic Unit (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 dichotomised the subjects into non-obese (BMI<30) and obese (BMI≥30) according to the WHO guidelines.

For beta diversity measurements, we used four distance measurements unweighted UniFrac, weighted UniFrac, generalised 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.²⁹ Different distance measurements represent different views of the microbial community and multiple distance measurements were 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).³⁰

In order to compare taxonomic abundance between groups, we used as input OTU (16S rRNA) and species (WGS) counts. Negative binomial regression was used

Table 1 Summary of demographic characteristics, sequencing methods and OTUs for included data sources

Database (Study)	Sample type	Sequencing method	Primers*	Species/OTU†	Sample source (N)	Sample size by group		Average BMI		P (obese vs non-obese)‡
						Obese	Non-obese	Obese	Non-obese	
Baxter <i>et al</i> ¹¹ (2016)	Stool	16S rRNA	V4	9997	Carcinoma (318); control (172)	118	368	34.05	24.93	<0.001
Feng <i>et al</i> ¹⁰ (2015)	Stool	WGS	NA	408 772	Adenoma (42); carcinoma (41); control (55)	43	113	31.98	25.62	<0.001
Vogtmann <i>et al</i> (2016)	Stool	WGS	NA	356 748	Carcinoma (52); control (52)	13	69	32.57	23.58	<0.001
Zackular <i>et al</i> ⁶ (2014)	Stool	16S rRNA	V4	24 990	Adenoma (30); carcinoma (30); control (30)	26	56	33.84	25.09	0.002
Zeller <i>et al</i> (2014a)	Stool	16S rRNA	V4	9969	Control (75); CRC (41)	19	108	32.53	24.13	<0.001
Zeller <i>et al</i> (2014b)	Tissue	16S rRNA	V4/V3-4	9988	Carcinoma (48); carcinoma-adjacent (48)	14	80	33.3	24.5	<0.001
Zeller <i>et al</i> (2014c)	Stool	WGS	NA	327 491	Adenoma (42); carcinoma (53); control (297)	34	160	32.15	24.18	<0.001

*1 Primers are NA for WGS.

†Average number of observed taxonomic 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.

BMI, body mass index; WGS, whole genome sequencing.

with BMI as a continuous variable for analysis of the microbiome while controlling for age and sex. Using multilevel modelling, the effects of confounders in study designs were examined. Details of all taxonomic analyses including random forest and mediation are found in online supplementary material 1.

RESULTS

Database and study selection

We performed a systematic review and meta-analysis guided search of the literature. Within this search we included five studies that analysed the microbiome of the stool or tissue from patients with colon cancer and which also had clinical information from patients on BMI (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.¹¹ The remaining studies were used as comparators to support or negate any associations found between the microbiome and obesity.

Characterisation of cofounders between studies

A major issue facing microbiome studies is the lack of standardised methods for collection, storage, nucleotide extraction, sequencing methodology and bioinformatic analysis. Thus, we began our analysis by characterising the effect of 16S rRNA variable region and sequencing methods (16S rRNA or WGS) on observed OTUs and Shannon diversity for 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 modelling to predict CRC status, we calculated the average log₂ OR (logit) for each study when these level 2 predictors (variable region and sequencing method) were 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; online supplementary figure S1). 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 predictive ability nor does variable region choice (figure 1A,B). Overall, among the potential cofounders 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.

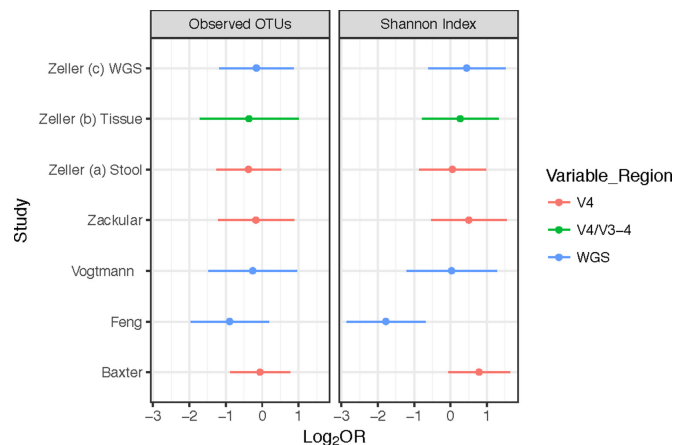


Figure 1 Variance in ability of alpha diversity to predict odds (log₂) of CRC controlling for obesity and study cofounders. The log₂ OR of CRC using observed OTUs (left panel) or Shannon Index (right panel) as predictors. The multilevel model includes obesity (level 1) and sequencing method (16S rRNA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

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 analyse alpha diversity within each sample study population, we calculated both richness, observed OTUs and Shannon diversity. We conducted linear modelling 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,²⁵ we also found significantly lower Shannon diversity in individuals that are obese without CRC 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 (figure 2A; online supplementary figure 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 *et al* dataset, which demonstrated a significantly higher alpha diversity with higher BMI both as continuous and categorical models, but not in the unadjusted analysis (figure 2B; online supplementary figure S2B and table 2). These results may be due to geography and diet of Asian populations. 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

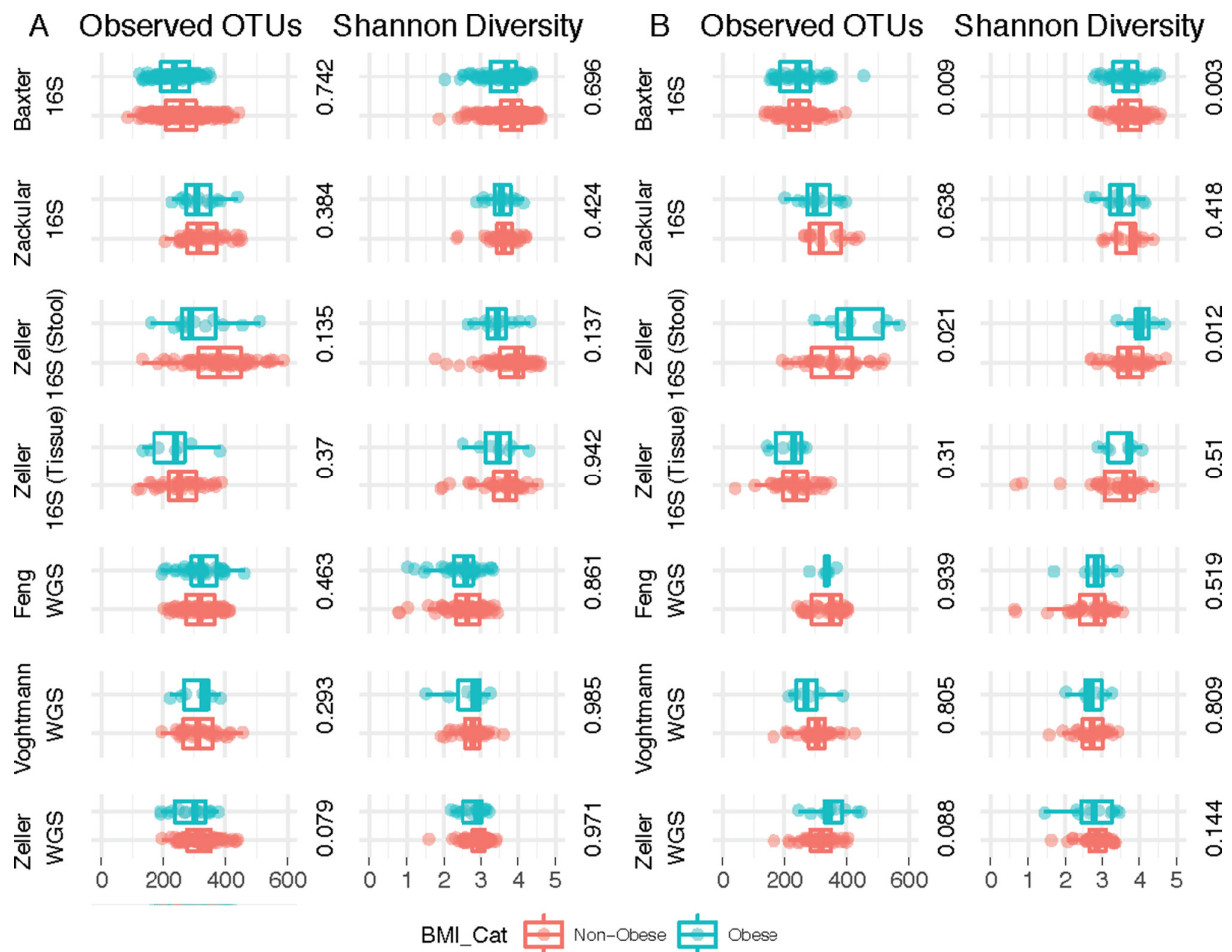


Figure 2 Alpha diversity in individuals with or without obesity and with or without CRC. (A) Observed OTUs and Shannon diversity in individuals without CRC or (B) with CRC comparing individuals with or without obesity. Reporting p values are from Mann-Whitney U test comparing the alpha diversity of individuals with or without obesity. BMI, body mass index; CRC, colorectal cancer; WGS, whole genome sequencing.

and normal weight individuals with or without CRC. To conduct this analysis, we calculated the omnibus p value for comparison of all distance matrices.³¹ In all of the data sets analysed, except Vogtman *et al* (WGS), Zeller *et al* (WGS) and Zeller *et al* (16S rRNA/tissue), we observed a significant difference (omnibus $p < 0.05$) in community structure between obese and non-obese individuals without CRC (table 3; see online supplementary figure S3A). This same analysis in individuals with CRC (obese vs non-obese), however, yielded only one significant observation in the Feng *et al* dataset (table 3; online supplementary figure 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.¹¹ From this analysis, controlling for age and sex, a significantly

lower relative abundance of several *Ruminococcus* 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*, Vogtman *et al*, Zeller *et al*, [WGS]), *Bifidobacterium* spp. (Zeller *et al* [WGS]) and *Akkermansia muciniphila* (Zackular *et al*, Zeller *et al* [WGS]) (online supplementary figure S4 and online supplementary table 1). When combining all differentially significant species, those from genus *Bacteroides* and *Bifidobacteria* appeared most often to differentiate individuals with and without obesity (online supplementary table 1). While no one genera or species was found to be differentially abundant (higher or lower) between all five 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 (online supplementary table 1).

**Table 2** Alpha-diversity analysis comparing obese and non-obese within cases and controls

Type	Reference (Sample)	Observed OTUs		Shannon	
		Linear (Est.)*	P value	Linear (Est.)*	P value
16S rRNA	Baxter et al¹¹ (Stool)				
	<i>Non-CRC-Ob vs Non-Ob</i>	-22.756	0.064	-0.173	0.046
	<i>CRC-Ob vs CRC Non-Ob</i>	-1.283	0.930	-0.059	0.564
	Zackular et al⁶ (Stool)				
	<i>Non-CRC-Ob vs Non-Ob</i>	-36.737	0.248	-0.240	0.217
	<i>CRC-Ob vs CRC Non-Ob</i>	7.006	0.811	0.053	0.794
	Zeller et al (Stool)				
	<i>Non-CRC - Ob v Non-Ob</i>	-86.434	0.049	-0.455	0.051
	<i>CRC - Ob v CRC Non-Ob</i>	-42.696	0.551	-0.020	0.953
	Zeller et al (Tissue)				
	<i>Non-CRC-Ob vs Non-Ob</i>	-14.373	0.682	-0.019	0.949
	<i>CRC-Ob vs CRC Non-Ob</i>	-12.389	0.683	0.140	0.690
WGS	Vogtmann et al (Stool)				
	<i>Non-CRC-Ob vs Non-Ob</i>	-25.023	0.399	-0.165	0.372
	<i>CRC-Ob vs CRC Non-Ob</i>	-9.187	0.698	-0.025	8.971
	Feng et al¹⁰ (Stool)				
	<i>Non-CRC-Ob vs Non-Ob</i>	-27.211	0.204	-0.327	0.085
	<i>CRC-Ob vs CRC Non-Ob</i>	21.947	0.348	0.389	0.183
	Zeller et al (Stool)				
	<i>Non-CRC-Ob vs Non-Ob</i>	-21.877	0.157	-0.219	0.041
	<i>CRC-Ob vs CRC Non-Ob</i>	3.592	0.832	-0.031	0.813

Values in bold are significant at $p < 0.05$.

*Linear modelling with BMI as continuous controlling for age and sex.

BMI, body mass index; CRC, colorectal cancer; WGS, whole genome sequencing.

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 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 (online supplementary table 2) and the overall presence or absence of each taxa in cases and controls (online supplementary table 3). Species from the *Bacteroides*, *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 = $-9e-05$ to -0.01) (online supplementary table 2), with the majority showing a negative effect and only 8/34 showing a positive effect; none showed a significant mediation effect (online supplementary table 2).

In our second approach, we derived an overall mediation effect using the FDR adjusted p values (q values); the q values were calculated for each data set and q values for taxa $< 20\%$ were considered to have a significant mediating effect (see online supplementary material 1). 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 (online supplementary 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*

Table 3 Beta-diversity analysis comparing obese and non-obese within cases and controls

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	F	P value
16S rRNA	Baxter et al¹¹ (Stool)												
	Non-CRC-Ob vs Non-Ob	1.852	0.011	1.871	0.027	1.798	0.111	1.792	0.019				0.024
	CRC-Ob vs Non-Ob	1.156	0.175	1.221	0.179	0.897	0.422	1.452	0.069				0.136
Zackular et al⁶ (Stool)	Non-CRC-Ob vs Non-Ob	1.314	0.088	1.784	0.022	2.824	0.007	1.494	0.054				0.015
	CRC-Ob vs 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 vs Non-Ob	1.691	0.006	1.565	0.040	1.387	0.180	1.393	0.101				0.014	
CRC-Ob vs 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 vs Non-Ob	1.159	0.189	1.411	0.060	1.596	0.095	1.597	0.027				0.066
	CRC-Ob vs Non-Ob	1.353	0.056	1.125	0.275	0.880	0.548	1.106	0.307				0.136
	Vogtmann et al (Stool)												
Non-CRC-Ob vs Non-Ob							0.897	0.485	1.041	0.422		0.508	
CRC-Ob vs Non-Ob							1.086	0.357	1.068	0.355		0.426	
Feng et al¹⁰ (Stool)	Non-CRC-Ob vs Non-Ob							3.129	0.003	4.144	0.008	0.004	
	CRC-Ob vs Non-Ob							2.923	0.019	4.972	0.008	0.010	
	Zeller et al (Stool)												
Non-CRC-Ob vs Non-Ob							1.821	0.081	2.892	0.047		0.065	
CRC-Ob vs Non-Ob							1.025	0.382	0.972	0.424		0.461	

F values from PERMANOVA testing on single or multiple distance metrics (omnibus test). Values in bold are significant at $p < 0.05$. CRC, colorectal cancer.



or *A. muciniphila*, were altered in individuals with obesity in their ability to differentiate CRC from non-CRC, we calculated the log₂ ORs for each species (online supplementary figure 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 (AUC) using age and sex as covariates 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 (figure 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 (figure 3A), demonstrating high heterogeneity among studies in predicting CRC or adenomas in obese individuals. Last, 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. 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 with 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.

Finally, we interrogated the metabolic potential of the bacterial community using the bioinformatics tool PICRUSt in order to obtain predicted functions. However, none of these predicted functions differentiated obese individuals among all studies (figure 4A,B). 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.^{32–41} In vivo, multiple studies indicate an interaction or mediating effect of the microbiome in promoting colon

tumourigenesis in the presence of a high-fat diet or genetic-induced obesity.^{18 19 42–44} 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-associated 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. Last, we show the microbiome, by itself or modelled 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.⁴⁵ 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 with 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 tumour location (left vs right), mutation profile, differentiation, mismatch repair status and diet, many of which are able to differentiate individuals with obesity among those with CRC.^{46–49} A high fat diet, however, 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 tumourigenesis from 30% to 60%.²⁸ Moreover, when faeces from high fat fed mice (K-ras^{G12Dint}) are transferred to healthy (K-ras^{G12Dint}) mice, tumour burden is increased along with diminished immune cell recruitment.²⁸ This was prevented, however, when supplemented with butyrate, which also increased *Bifidobacterium* abundance as compared with mice not supplemented with butyrate.²⁸ 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 fibre, low-fat rural African diet.⁵⁰ 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

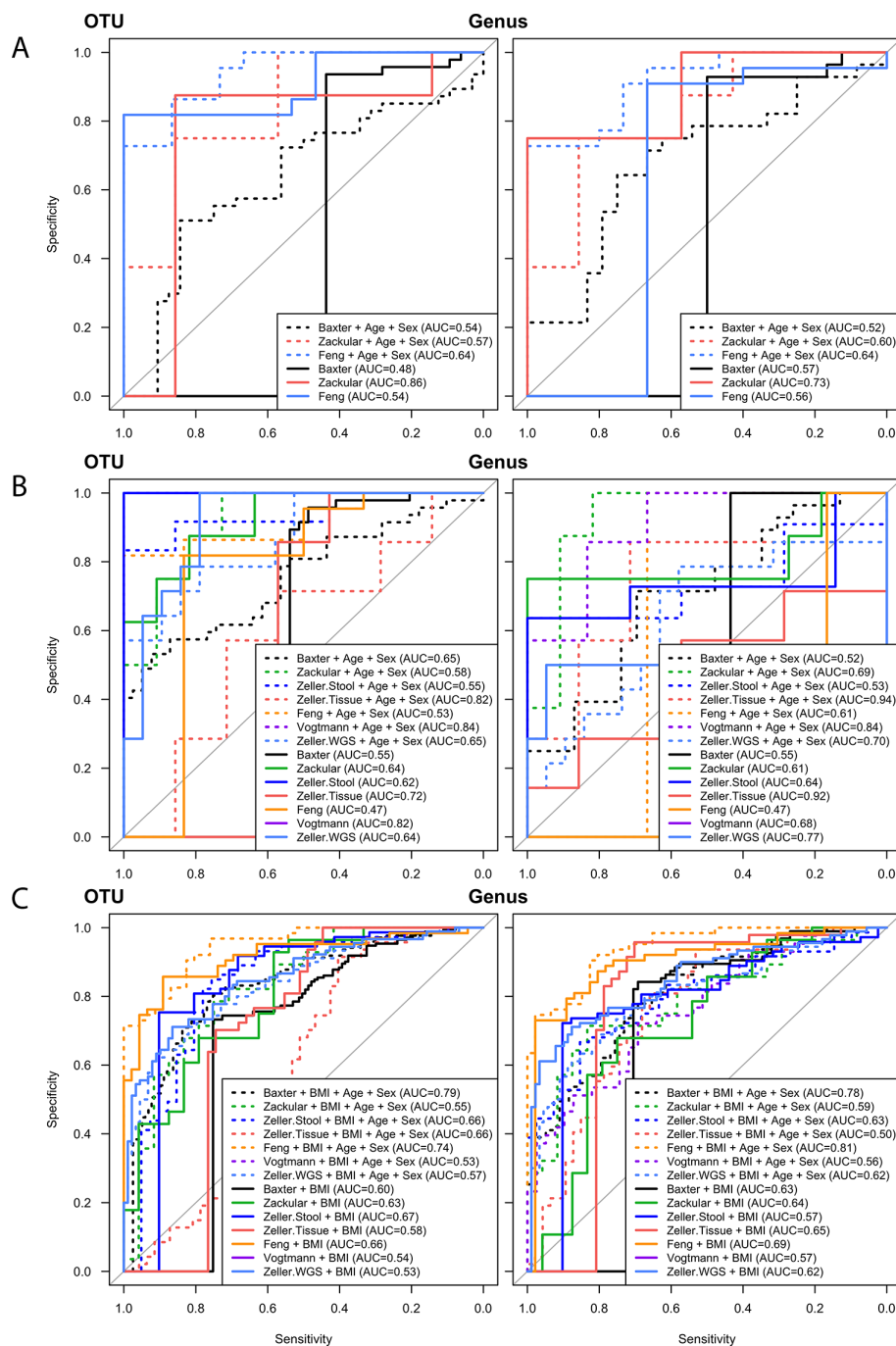


Figure 3 Microbial classifiers of CRC and obesity-associated CRC. (A) ROC for the random forest classification analyses for obese vs non-obese in individuals with CRC for each study. AUC is the 10-fold cross validated area under the curve. (B) ROC for the random forest classification analyses of obese 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. AUC, area under the receiver operating curve; BMI, body mass index; CRC, colorectal cancer; ROC, receiver operating curve; WGS, whole genome sequencing.

a proinflammatory environment conducive to colon carcinogenesis.

BMI is crude measure of obesity, and other more accurate measures (eg, waist circumference, adipokines and so on) 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 that 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.⁵¹ Thus, this study sets the stage for future research to consider adding measures of adiposity beyond BMI

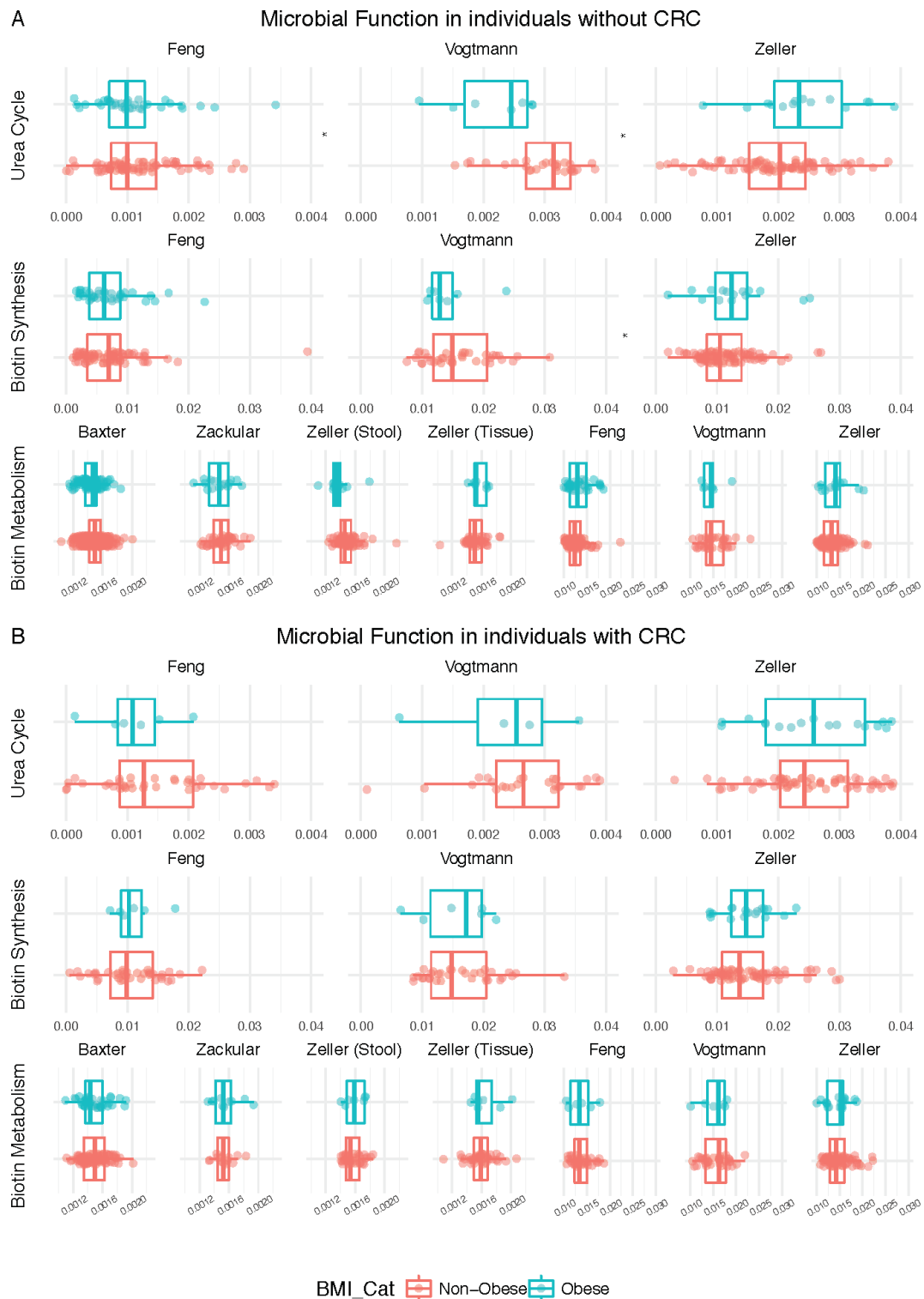


Figure 4 Pathway abundance analysis in individuals with or without obesity among individuals with or without CRC. relative abundance of KEGG metabolic pathways (16S rRNA) or modules (WGS) inferred from PICRUSt or HUMAnN, respectively. Significance was calculated using the Wilcox test correction for multiple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

when studying the aetiology and risk of CRC as well as other cancers influenced by obesity.

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, while this study did not assess

diet directly, other studies suggest that diet, rather than obesity, is creating a proinflammatory microbial community increasing CRC risk.^{28 50} Hence, characterising the role of the diet in addition to the microbiome in CRC aetiology is necessary, which will require more detailed molecular analyses and well-designed longitudinal

human studies to identify dietary and early stage microbial biomarkers prior to disease.

Author affiliations

¹Nutrition Sciences Division, Robbins College of Health and Human Science, Baylor University, Waco, Texas, USA

²Department of Biology, Baylor University, Waco, Texas, USA

³Respera Biosciences, Baltimore, Maryland, USA

⁴Department of Educational Psychology, Baylor University, Waco, Texas, USA

⁵Department of Surgery, Mayo Clinic, Rochester, New York, USA

⁶Department of Physiology and Biomedical Engineering, Mayo Clinic, Rochester, New York, USA

⁷Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, New York, USA

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 reproducibility. Our thanks to the members of the Baylor Writing Group, Joe Taube, Karen Melton, Elise King, Elyssia Gallagher, for their feedback and editing.

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

Funding This work was supported by the Baylor University Summer Sabbatical Grant (PI—summer salary support). NC was supported by the NCI award R01CA179243.

Competing interests JRW is a significant shareholder in the company Respera Biosciences LLC.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement All data are publicly available and all processed data are available at https://github.com/GreathouseLab/CRC_BMI_meta_analysis

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- Arnold M, Pandeya N, Byrnes G, *et al.* Global burden of cancer attributable to high body-mass index in 2012: a population-based study. *Lancet Oncol* 2015;16:36–46.
- Whiteman DC, Wilson LF. The fractions of cancer attributable to modifiable factors: a global review. *Cancer Epidemiol* 2016;44:203–21.
- Giovannucci E, Ascherio A, Rimm EB, *et al.* Physical activity, obesity, and risk for colon cancer and adenoma in men. *Ann Intern Med* 1995;122:327–34.
- Engeland A, Tretli S, Austad G, *et al.* Height and body mass index in relation to colorectal and gallbladder cancer in two million Norwegian men and women. *Cancer Causes Control* 2005;16:987–96.
- Calle EE, Kaaks R, Overweight KR. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. *Nat Rev Cancer* 2004;4:579–91.
- Zackular JP, Baxter NT, Iverson KD, *et al.* The gut microbiome modulates colon tumorigenesis. *MBio* 2013;4:e00692–13.
- Zackular JP, Baxter NT, Chen GY. *Manipulation of the gut microbiota reveals role in colon tumorigenesis*. 1. mSphere, 2016.
- Bullman S, Pedamallu CS, Sicinska E, *et al.* Analysis of *Fusobacterium* persistence and antibiotic response in colorectal cancer. *Science* 2017;358:1443–8.
- Chung L, Thiele Orberg E, Geis AL, *et al.* *Bacteroides fragilis* toxin coordinates a pro-carcinogenic inflammatory cascade via targeting of colonic epithelial cells. *Cell Host Microbe* 2018;23:203–14.
- Feng Q, Liang S, Jia H, *et al.* Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat Commun* 2015;6:6528.
- Baxter NT, Ruffin MT, Rogers MA, *et al.* Microbiota-Based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. *Genome Med* 2016;8:37.
- Nakatsu G, Li X, Zhou H, *et al.* Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nat Commun* 2015;6:8727.
- Dejea CM, Fathi P, Craig JM, *et al.* Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science* 2018;359:592–7.
- Purcell RV, Pearson J, Aitchison A, *et al.* Colonization with enterotoxigenic *Bacteroides fragilis* is associated with early-stage colorectal neoplasia. *PLoS One* 2017;12:e0171602.
- Yang Y, Weng W, Peng J, *et al.* *Fusobacterium nucleatum* increases proliferation of colorectal cancer cells and tumor development in mice by activating Toll-like receptor 4 signaling to nuclear factor- κ B, and up-regulating expression of microRNA-21. *Gastroenterology* 2017;152:e24:851–66.
- Everard A, Lazarevic V, Gaïa N, *et al.* Microbiome of prebiotic-treated mice reveals novel targets involved in host response during obesity. *ISME J* 2014;8:2116–30.
- Ley RE, Turnbaugh PJ, Klein S, *et al.* Microbial ecology: human gut microbes associated with obesity. *Nature* 2006;444:1022–3.
- Pfalzer AC, Kamanu FK, Parnell LD, *et al.* Interactions between the colonic transcriptome, metabolome, and microbiome in mouse models of obesity-induced intestinal cancer. *Physiol Genomics* 2016;48:545–53.
- Pfalzer AC, Nesbeth PD, Parnell LD, *et al.* Diet- and Genetically-Induced obesity differentially affect the fecal microbiome and metabolome in *Apc1638N* mice. *PLoS One* 2015;10:e0135758.
- Turnbaugh PJ, Bäckhed F, Fulton L, *et al.* Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* 2008;3:213–23.
- Turnbaugh PJ, Hamady M, Yatsunenko T, *et al.* A core gut microbiome in obese and lean twins. *Nature* 2009;457:480–4.
- Turnbaugh PJ, Ley RE, Mahowald MA, *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 2006;444:1027–31.
- Ridaura VK, Faith JJ, Rey FE, *et al.* Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* 2013;341:1241214.
- Walters WA, Xu Z, Knight R. Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS Lett* 2014;588:4223–33.
- Sze MA, Schloss PD. Looking for a signal in the noise: revisiting obesity and the microbiome. *MBio* 2016;7:e01018-16.
- Caër C, Rouault C, Le Roy T, *et al.* Immune cell-derived cytokines contribute to obesity-related inflammation, fibrogenesis and metabolic deregulation in human adipose tissue. *Sci Rep* 2017;7:3000.
- Chehimi M, Vidal H, Eljaafari A. Pathogenic role of IL-17-producing immune cells in obesity, and related inflammatory diseases. *J Clin Med* 2017;6:68.
- Schulz MD, Atay C, Heringer J, *et al.* High-fat-diet-mediated dysbiosis promotes intestinal carcinogenesis independently of obesity. *Nature* 2014;514:508–12.
- Arumugam M, Raes J, Pelletier E, *et al.* Enterotypes of the human gut microbiome. *Nature* 2011;473:174–80.
- Chen J, Bittinger K, Charlson ES, *et al.* Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics* 2012;28:2106–13.
- Chen J, King E, Deek R, *et al.* An omnibus test for differential distribution analysis of microbiome sequencing data. *Bioinformatics* 2018;34:643–51.
- Kellermayer R, Dowd SE, Harris RA, *et al.* Colonic mucosal DNA methylation, immune response, and microbiome patterns in Toll-like receptor 2-knockout mice. *Faseb J* 2011;25:1449–60.
- Everard A, Belzer C, Geurts L, *et al.* Cross-talk between *Akkermansia muciniphila* and intestinal epithelium controls diet-induced obesity. *Proceedings of the National Academy of Sciences* 2013;110:9066–71.
- Garidou L, Pomié C, Klopp P, *et al.* The Gut Microbiota Regulates Intestinal CD4 T Cells Expressing ROR γ t and Controls Metabolic Disease. *Cell Metab* 2015;22:100–12.
- Moya-Pérez A, Neef A, Sanz Y. 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* 2015;10:e0126976.
- Schneeberger M, Everard A, Gómez-Valadés AG, *et al.* *Akkermansia muciniphila* inversely correlates with the onset of inflammation,



- altered adipose tissue metabolism and metabolic disorders during obesity in mice. *Sci Rep* 2015;5:16643.
37. Greer RL, Dong X, Moraes ACF, *et al.* Akkermansia muciniphila mediates negative effects of IFN γ on glucose metabolism. *Nat Commun* 2016;7:13329.
 38. Zeng H, Ishaq SL, Zhao FQ, *et al.* Colonic inflammation accompanies an increase of β -catenin signaling and Lachnospiraceae/Streptococcaceae bacteria in the hind gut of high-fat diet-fed mice. *J Nutr Biochem* 2016;35:30–6.
 39. Zhao S, Liu W, Wang J, *et al.* Akkermansia muciniphila improves metabolic profiles by reducing inflammation in Chow diet-fed mice. *J Mol Endocrinol* 2017;58:1–14.
 40. Thaiss CA, Levy M, Grosheva I, *et al.* Hyperglycemia drives intestinal barrier dysfunction and risk for enteric infection. *Science* 2018;359:1376–83.
 41. Liu R, Hong J, Xu X, *et al.* Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention. *Nat Med* 2017;23:859–68.
 42. Higashimura Y, Naito Y, Takagi T, *et al.* Protective effect of agaro-oligosaccharides on gut dysbiosis and colon tumorigenesis in high-fat diet-fed mice. *Am J Physiol Gastrointest Liver Physiol* 2016;310:G367–G375.
 43. Liu W, Crott JW, Lyu L, *et al.* 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* 2016;7:1780–90.
 44. Koh GY, Kane A, Lee K, *et al.* 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* 2018.
 45. Sze MA, Schloss PD. Leveraging existing 16S rRNA gene surveys to identify reproducible biomarkers in individuals with colorectal tumors. *MBio* 2018;9.
 46. Hanyuda A, Cao Y, Hamada T, *et al.* Body mass index and risk of colorectal carcinoma subtypes classified by tumor differentiation status. *Eur J Epidemiol* 2017;32:393–407.
 47. Liu L, Tabung FK, Zhang X, *et al.* Diets that promote colon inflammation associate with risk of colorectal carcinomas that contain Fusobacterium nucleatum. *Clin Gastroenterol Hepatol* 2018;16:1622–31.
 48. Hanyuda A, Ogino S, Qian ZR, *et al.* Body mass index and risk of colorectal cancer according to tumor lymphocytic infiltrate. *Int J Cancer* 2016;139:854–68.
 49. Sinicrope FA, Foster NR, Yoon HH, *et al.* 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. *JCO* 2012;30:406–12.
 50. O'Keefe SJ, Li JV, Lahti L, *et al.* Fat, fibre and cancer risk in African Americans and rural Africans. *Nat Commun* 2015;6:6342.
 51. Yu D, Zheng W, Johansson M, *et al.* Overall and central obesity and risk of lung cancer: a pooled analysis. *J Natl Cancer Inst* 2018;110:831–42.