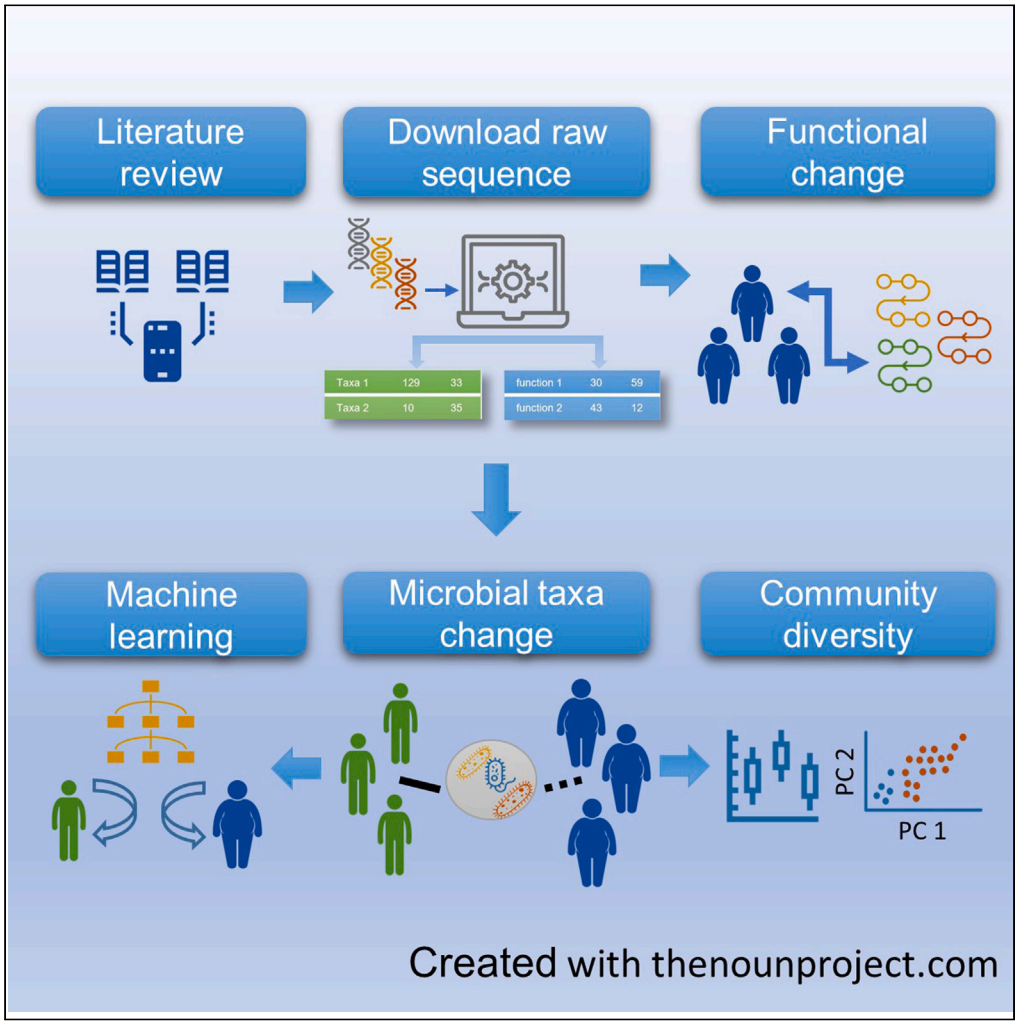
iScience

CellPress
OPEN ACCESS

Article

Combing fecal microbial community data to identify consistent obesity-specific microbial signatures and shared metabolic pathways



Yu Lin, Zhilu Xu, Yun Kit Yeoh, ..., Wei Jiang, Francis Ka Leung Chan, Siew Chien Ng

siewchienng@cuhk.edu.hk

Highlights

A decreased microbial diversity was identified in obesity across 18 study cohorts

Microbiome analysis reveals common obesity-associated depleted taxa

Functional alteration of obesity suggests metabolic adaptation to a high-fat diet

Lin et al., iScience 26, 106476
April 21, 2023
https://doi.org/10.1016/j.isci.2023.106476

iScience

CellPress
OPEN ACCESS

Article

Combing fecal microbial community data to identify consistent obesity-specific microbial signatures and shared metabolic pathways

Yu Lin,[1,2,3] Zhilu Xu,[1,2,3] Yun Kit Yeoh,[1,2,4] Hein Min Tun,[1,5] Wenli Huang,[1,2,3] Wei Jiang,[1,2,3] Francis Ka Leung Chan,[1,2,3] and Siew Chien Ng[1,2,3,6,*]

SUMMARY

Obesity is associated with altered gut microbiome composition but data across different populations remain inconsistent. We meta-analyzed publicly available 16S-rRNA sequence datasets from 18 different studies and identified differentially abundant taxa and functional pathways of the obese gut microbiome. Most differentially abundant genera (*Odoribacter*, *Oscillospira*, *Akkermansia*, *Alistipes*, and *Bacteroides*) were depleted in obesity, indicating a deficiency of commensal microbes in the obese gut microbiome. From microbiome functional pathways, elevated lipid biosynthesis and depleted carbohydrate and protein degradation suggested metabolic adaptation to high-fat, low-carbohydrate, and low-protein diets in obese individuals. Machine learning models trained on the 18 studies were modest in predicting obesity with a median AUC of 0.608 using 10-fold cross-validation. The median AUC increased to 0.771 when models were trained in eight studies designed for investigating obesity-microbiome association. By meta-analyzing obesity-associated microbiota signatures, we identified obesity-associated depleted taxa that may be exploited to mitigate obesity and related metabolic diseases.

INTRODUCTION

Obesity is a growing epidemic worldwide and is associated with increased morbidity and mortality.[1] The obesity crisis is the result of a combination of environmental and personal factors, energy balance dysregulation, host genetics, and gut microbiota changes.[2] Several lines of evidence indicate that the gut microbiome is linked to obesity. In studies of mice and human twin pairs, altered core gut microbiome composition was associated with obesity.[3,4] Moreover, a higher relative abundance of *Firmicutes* in obese individuals was shown to harvest more energy from diet.[3,4] In addition, stool samples from obese monozygotic twin pairs, when implanted into germ-free mice, resulted in weight gain whereas mice receiving stools from lean twins maintained normal weight.[5] Reproducible evidence from several clinical trials reported increased levels of insulin sensitivity and butyrate-producing bacteria in patients with obesity or metabolic diseases following fecal microbiota transplant (FMT) from lean donors.[6–8] The change in gut microbiome composition in obese individuals was associated with up-regulated gene expression of fasting-induced adipose factor in the colon, which may impair triglyceride metabolism and induce fat storage.[9] Collectively, existing knowledge supports the mechanistic involvements of gut microbiota in obesity and suggests that using gut microbiome-based therapeutics to manage obesity has great potential.

As such, there is a growing interest in identifying gut microbial taxa that can be exploited to modulate obesity. A human gut microbiome survey found that *Actinobacteria* and several genera in the *Firmicutes* phylum able to harvest energy from carbohydrates were positively correlated with body mass index (BMI). A higher *Firmicutes* to *Bacteroidetes* (F/B) ratio was reported to be indicative of obesity,[10] although similar findings were not reproduced in other human gut microbiota surveys.[11,12] Several bacterial genera such as *Alistipes*, *Akkermansia*, and *Lactobacillus*[13–16] have been consistently reported to be altered in obese individuals. However, other species had shown inconsistent associations across different studies. For example, several studies reported increased abundances of *Bacteroides*, *Bifidobacterium*, and *Prevotella* in obese individuals,[14,17,18] whereas others report no significant associations or opposite trends.[13,19]

[1]Microbiota I-Center (MagIC), Hong Kong SAR, China

[2]Center for Gut Microbiota Research, Faculty of Medicine, the Chinese University of Hong Kong, Hong Kong SAR, China

[3]Department of Medicine and Therapeutics, Institute of Digestive Disease, Li Ka Shing Institute of Health Sciences, Faculty of Medicine, the Chinese University of Hong Kong, Hong Kong SAR, China

[4]Department of Microbiology, Faculty of Medicine, the Chinese University of Hong Kong, Hong Kong SAR, China

[5]Jockey Club School of Public Health and Primary Care, Faculty of Medicine, the Chinese University of Hong Kong, Hong Kong SAR, China

[6]Lead contact

*Correspondence: siewchienng@cuhk.edu.hk

https://doi.org/10.1016/j.isci.2023.106476

Table 1. Characteristics of studies included in the meta-analysis

Study	Sample size	Obese subjects	Obese percentage amongst total	Population	Geographical region	Research design
Ahmad[23]	60	40	66.67%	Adult	Pakistan	[a]
Chavez[24]	67	42	62.69%	Adult	Mexico	[a]
Barengolts[25]	89	49	55.06%	Adult	USA	[a]
Gao[26]	77	39	50.65%	Children	China	[a]
Olsson[27]	41	17	41.46%	Adult	France	[a]
Lippert[28]	20	11	55.00%	Adult	Austria	[a]
Ross[29]	63	38	60.32%	Adult	USA	[a]
Zupancic[30]	168	61	36.31%	Adult	USA	[a]
Escobar[31]	17	8	47.06%	Adult	Colombia	[b]
Wu[32]	43	5	11.63%	Adult	USA	[b]
Zeevi[33]	418	135	32.30%	Adult	Israel	[b]
Goodrich[34]	644	193	29.97%	Adult	Britain	[b]
Baxter[35]	115	47	40.87%	Adult	USA/Canada	[b]
HKGutMicMap[36]	226	48	21.24%	Adult	China	[b]
He[37]	2872	1040	36.21%	Adult	China	[b]
HMP[38]	151	23	15.23%	Adult	USA	[b]
AGP[39]	2204	278	12.61%	Adult	USA/Britain/Canada	[b]
Stanislawski[40]	119	24	20.17%	Children	USA	[b]

[a]Studies designed for investigating obesity-microbiome association.
[b]Studies that included general population with BMI information.

Given the complexity of the gut ecosystem, microbial taxa that are consistently connected to obesity are not well defined. By combining the multi-cohort sequence data, a meta-analysis with a larger sample size than an individual study can increase statistical power and identify common signatures across several datasets to derive a consensus.[20]

Several meta-analyses have investigated the association between obesity and gut microbiota. Sze et al.[21] conducted a meta-analysis of 10 studies and found decreased alpha diversity in obese individuals. Furthermore, their reported random forest model for obesity classification performed suboptimally in distinguishing obese and healthy subjects, with a median accuracy of 0.56. Stanislawski et al.[22] included five study cohorts with diverse racial backgrounds and found that associations between microbial diversity and obesity were ethnicity-specific. Alpha diversity was negatively associated with BMI in Caucasian populations and the *Prevotella* genus was positively correlated with BMI in African populations. However, the specific role of microbial taxa associated with obesity has yet to be clearly established. Moreover, microbial functional changes have not been investigated in prior meta-analyses. Here, we conduct a comprehensive meta-analysis incorporating newly published studies to identify the effects of obesity on the gut microbiota composition and functional pathway changes.

RESULTS

Literature search and study characteristics

We retrieved 2,568 studies according to the search strategy (Figure S1). The reference lists of two published meta-analyses related to obesity[21,22] were also referenced as an extended search. In total, we identified 2,583 studies for further selection. The study selection process is shown in Figure S2. Articles without full-text and those that did not fulfill our study goals were removed by examining the title and abstract (remaining n = 2508). Then, we excluded studies that did not meet predefined criteria (remaining n = 29). All relevant studies were screened by four authors. After identifying the availability of sequence and metadata in the remaining 29 studies, we finally included 18 studies for meta-analysis (Tables 1 and S1).[23–40] Most of these studies focused on adult obesity and their subjects represented various ethnic backgrounds across the world. Among these 18 studies, seven of them were designed for investigating associations between

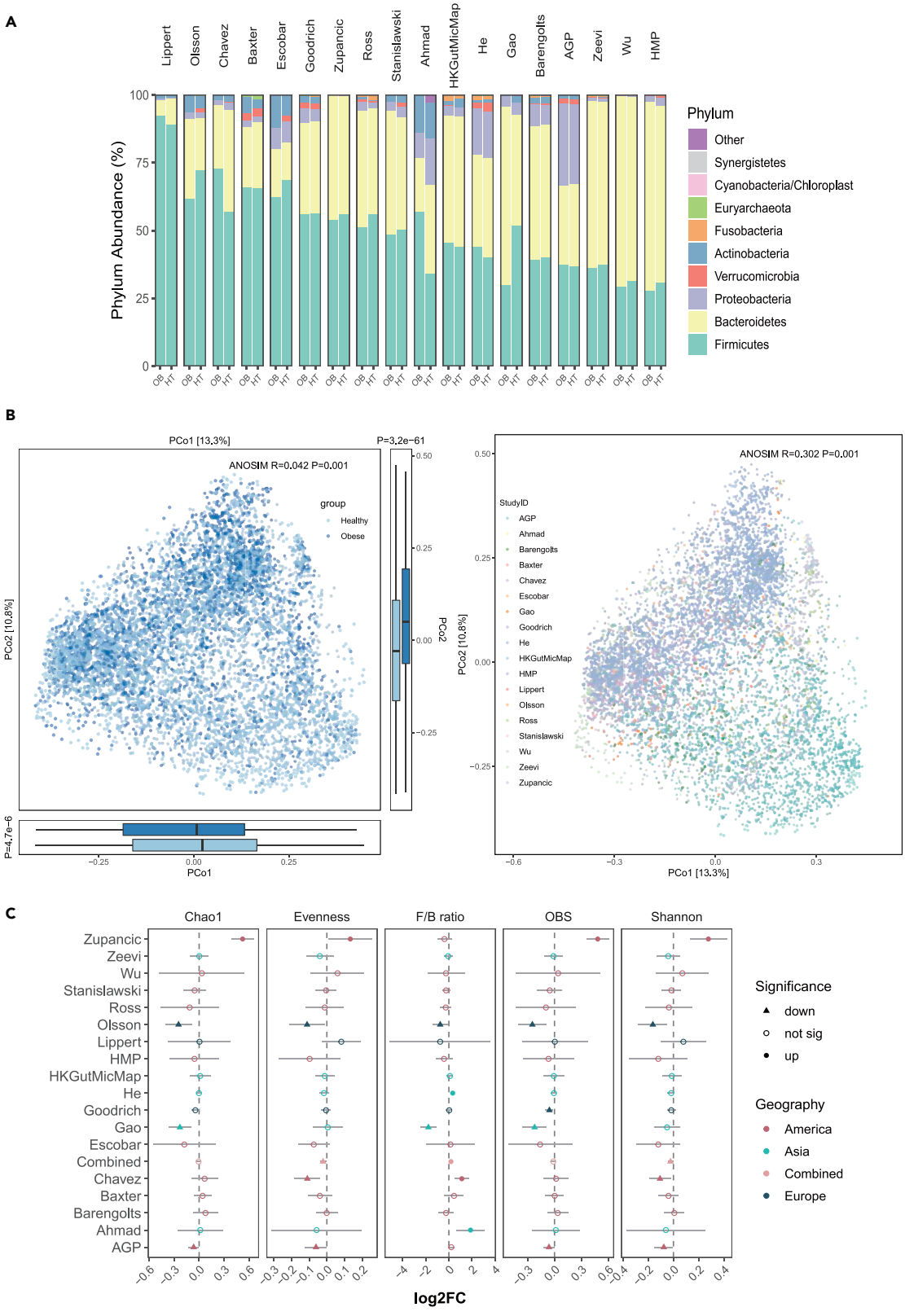

A
Lippert
Olsson
Chavez
Baxter
Escobar
Goodrich
Zupancic
Ross
Stanislawski
Ahmad
HKGutMicMap
He
Gao
Barengolts
AGP
Zeevi
Wu
HMP
Phylum Abundance (%)
100
75
50
25
0
OB
HT
Phylum
Other
Synergistetes
Cyanobacteria/Chloroplast
Euryarchaeota
Fusobacteria
Actinobacteria
Verrucomicrobia
Proteobacteria
Bacteroidetes
Firmicutes
B
PCo1 [13.3%]
P=3.2e−61
ANOSIM R=0.042 P=0.001
group
Healthy
Obese
PCo2 [10.8%]
P=4.7e−6
PCo1
ANOSIM R=0.302 P=0.001
StudyID
AGP
Ahmad
Barengolts
Baxter
Chavez
Escobar
Gao
Goodrich
He
HKGutMicMap
HMP
Lippert
Olsson
Ross
Stanislawski
Wu
Zeevi
Zupancic
C
Chao1
Evenness
F/B ratio
OBS
Shannon
Zupancic
Zeevi
Wu
Stanislawski
Ross
Olsson
Lippert
HMP
HKGutMicMap
He
Goodrich
Gao
Escobar
Combined
Chavez
Baxter
Barengolts
Ahmad
AGP
log2FC
Significance
down
not sig
up
Geography
America
Asia
Combined
Europe

Figure 1. The effect of obesity on phylum level community structure, beta diversity, alpha diversity, and F/B ratio across 18 studies

(A) Stacked bar plot denotes the relative abundance of the top 10 phyla in the obese and healthy groups across different studies, ordered by the abundance of *Firmicutes* phylum in each cohort. OB: obese group, HT: healthy group.

(B) Left: The PCoA analysis based on the Bray-Cutis distance, showing the distance of the microbial taxa between obese and lean individuals. The bar plot at the bottom represents the difference of principal coordinates 1 (PC1) axis in the obese and healthy populations. The bar plot on the right represents the difference in the PC2 axis in the obese and healthy populations. The light blue bars represent healthy individuals, and the dark blue bars represent obese individuals. The p value within the boxplot was computed by the Wilcoxon rank-sum test between the obese and healthy groups. Right: The PCoA analysis based on the Bray-Cutis distance of bacterial communities among different studies. The ANOSIM test was used to identify the statistical difference between individuals with different obesity statuses and individuals from different studies based on the Bray-Cutis distance. The ANOSIM R indicates the dissimilarity among different groups (obese versus non-obese, different studies).

(C) The forest plot illustrates the $\log_2$ fold change($\log_2$FC) of the obese group over the healthy group in alpha diversity metrics including Shannon, Evenness, Observed OTUs (OBS), Chao1 index, and the F/B ratio. The studies are organized by geographic region. The gray line indicates the confidence interval of the $\log_2$ fold change, identified by Welch's t-test. The significance of $\log_2$ fold change is denoted by filled/unfilled circles. The combined $\log_2$ fold change is calculated by the random effects model. When the $\log_2$ fold change is larger than 0, obese individuals have a higher level of measurement than non-obese individuals. HMP: Human Microbiome Project, AGP: American Gut Project.

gut microbiota and obesity. After excluding overweight (but not obese) individuals, 7,394 individuals remained, comprising of 2,098 obese and 5,296 healthy lean subjects.

Gut microbial community diversity and obesity

The composition of the gut microbiota at the phylum level for each study is shown in Figure 1A. In all studies, *Firmicutes* and *Bacteroidetes* were consistently the two most dominant phyla. For beta diversity, Bray-Curtis dissimilarities were calculated for the combined studies based on amplicon sequence variants (ASVs) collapsed to the genus level and visualized by Principal Coordinate Analysis (PCoA). ASV represents the actual amplicon sequence from each 16S rRNA gene copy and is the basic unit for analyzing microbial communities. There was a significant difference for principal coordinates 1 (PCo1) and PCo2 between the obese and healthy groups ($p< 4.7e-6$ for PCo1 and $p< 3.2e-61$ for PCo2, Figure 1B), as well as indifferent studies ($p< 2.2e-16$ for both PCo1 and PCo2, Figures S4B and S4C). ANOSIM tests were applied to assess statistically significant differences in dissimilarities between the obese and healthy groups and across studies. With all 18 studies combined, the ANOSIM test (Figure 1B) revealed a statistically significant dissimilarity ($R = 0.042$, $p = 0.001$) between obese and lean individuals. For dissimilarity among cohorts from the 18 studies, the ANOSIM test (Figure 1B) was statistically significant ($p = 0.001$), with $R = 0.303$. The larger R value from comparing different studies versus that from comparing obese and healthy groups in all studies combined indicated that a larger proportion of variance in the combined dataset could be attributable to between-study variability rather than obesity status. The gut microbial community composition was more similar within each study, but different in comparison to obese individuals across other studies ($R = 0.303$ for variance contributed by study effect versus $R = 0.042$ for variance contributed by obesity). This observation reflects a hierarchical structure in the combined microbial data, and we therefore applied random effects models[41] to estimate the fixed effects of obesity on microbial community diversity and composition.

To evaluate associations between obesity and alpha diversity, $\log_2$ fold change for alpha diversity metrics including Shannon, Evenness, Chao1 index, Observed OTUs (OBS), and F/B ratio in the combined cohort was computed using a random effects model with adjustment for study effect. In the combined 18 studies, we observed statistically significant decreases in Shannon diversity ($p = 0.03$) and Evenness ($p = 0.01$) and significant increases in F/B ratio ($p< 0.001$) in obese subjects compared with healthy subjects (Figure 1C, Tables S2–S6). We applied I^2 statistics for study heterogeneity evaluation. The between-study heterogeneity variance (I^2(Shannon) = 51.2%, I^2(Evenness) = 41.9%, I^2(F/B ratio) = 75.7%) was moderate to substantial in 18 cohorts (Tables S2–S6). The highest study heterogeneity was contributed by Zupancic et al., Chavez et al., Gao et al., and Olsson et al., as shown in the Baujat plot (Figures S3A, S3C, and S3E). No significant publication bias was observed according to the funnel plot and Egger's regression test of funnel plot asymmetry (Figures S3B, S3D, S3F, Tables S2–S6). During the literature search, we found six included studies (Barengolts et al., Olsson et al., Ross et al., Goodrich et al., Baxter et al., and Stanislawski et al.) without medication history on antibiotics usage. To eliminate the potential impact of antibiotics on the microbiome composition, we conducted sensitivity analysis for the combined alpha diversity to determine if our findings were robust. We excluded six studies without antibiotics usage information and applied the random effects model to re-evaluate combined alpha diversity for obesity using the remaining 12 studies. The random effects model on combined alpha diversity of the remaining 12 studies showed consistent findings that a

significantly lower microbiota Evenness ($p = 0.01$) and a significantly increased F/B ratio ($p < 0.001$) in obese subjects, whereas a similar trend of combined Shannon diversity in obese subjects despite lack of statistical significance (Figure S4A and Table S7). The sensitivity analysis indicated that studies without antibiotics usage information did not largely influence the findings of the meta-analysis. As the meta-analysis published by Stanislawski et al.[22] reported that Shannon index may be positively correlated with BMI in Caucasian populations, we attempted to identify relationships between Shannon diversity and BMI using a random effects model in the combined cohort while accounting for racial ethnic groups. As five studies (Ahmad et al., Chavez et al., Gao et al., Olsson et al., and Lippert et al.) did not provide BMI, we combined the remaining 13 studies to assess correlations between BMI and gut microbial community diversity. Relationships between Shannon diversity, F/B ratio, and BMI are shown in Figure S5. Only F/B ratio with log transformation was positively associated with BMI in the Asian population. To minimize the influence of study variation, we selected only two studies (AGP and HMP) that encompass several racial ethnic backgrounds, including Caucasians, Asians, and Africans, to identify relationships between community diversity and BMI. We found that Shannon index and Evenness were negatively associated with BMI in the Caucasian population ($p < 0.05$, random effects model) (Figure S6). For the relationship between alpha diversity and BMI in populations stratified by age group and sex, only Evenness has a negative association with BMI in adult and female individuals (Figure S7).

Differentially abundant microbial genera in obese and healthy lean subjects

We collapsed ASVs according to their genus classifications because genus level units are more comparable across study cohorts and less sensitive to batch effects.[42] The distribution of genus level units across the 18 studies is visualized in the UpSet plot in Figure S8. We included 157 genera detected in over 2/3 of studies (12 or more studies) for further analysis. A random-effects model was applied to identify differentially abundant genera in obese and healthy subjects (Figure 2A). We also used Stratified Wilcoxon tests to assess the significance of these genera to corroborate the random effect model (Figure 2B). The Spearman coefficient indicated that the results of the two methods were positively correlated with each other. The 26 genera with p-value <0.05 (random effects model) and Stratified Wilcoxon FDR <0.001 were used as core taxa for further analysis. The differential abundance analysis (Figure 2C) revealed that five genera, *Allisonella*, *Agathobacter*, *Dorea*, *Negativibacillus*, and *Roseburia*, were enriched in obese individuals whereas the other 21 which included *Odoribacter*, *Oscillospira*, *Akkermansia*, *Alistipes*, and *Bacteroides*, were depleted in obese individuals (Tables S8 and S9). This analysis suggests that there is a distinct shift in gut microbiome composition in obese individuals relative to healthy controls.

Genus and species level microbial markers correlated with BMI

Using the 26 differentially abundant genera identified above, we next examined their relationship with BMI using random effects models. Among the 21 obesity-associated depleted genera, 12 were significantly correlated with BMI in the combined population (Figure 3A and Table S10). When controlling for ethnic background, 14 of the 21 genera were significantly associated with BMI in the Asian population. In the Caucasian and African populations, there was a negative trend between the selected genera and BMI, although these trends were not statistically significant. The relationship between obesity-associated enriched genera and BMI identified across 18 studies is shown in Figure S9 and Table S11. Among five obesity-associated enriched genera, we found that four, including *Allisonella*, *Dorea*, *Agathobacter*, and *Negativibacillus*, were positively correlated with BMI in the combined population. After stratification by ethnicity, *Allisonella*, *Dorea*, and *Agathobacter*, were significantly correlated with BMI in the Asian population and *Dorea*, *Agathobacter*, and *Negativibacillus* were significantly associated with BMI in the Caucasian population (Figure S9). From the subgroup analysis based on age, we found that some identified genus markers (11 of 21 depleted genera and 3 of 5 enriched genera in obese participants) were correlated with the BMI of adults (Figure S10). However, all taxa did not show significant associations with BMI in the child population. In individuals stratified by sex, *Allisonella* and *Negativibacillus* were positively associated with BMI in both males and females (Figure S10).

To identify species level markers associated with BMI, as opposed to genera units in 16S data, we inferred species-level community composition using metagenomic sequencing data from the HKGutMicMap cohort. There were 57 species from the HKGutMicMap cohort belonging to the 21 genera associated with BMI in the Asian population. Among the 57 species, only four (*Bacteroides nordii*, *Bacteroides thetaiotaomicron*, *Intestinimonas butyriciproducens*, and *Ruthenibacterium lactatiformans*) were negatively correlated with BMI according to Spearman correlation (Figure 3B, Figure S11A, and Table S12). For species

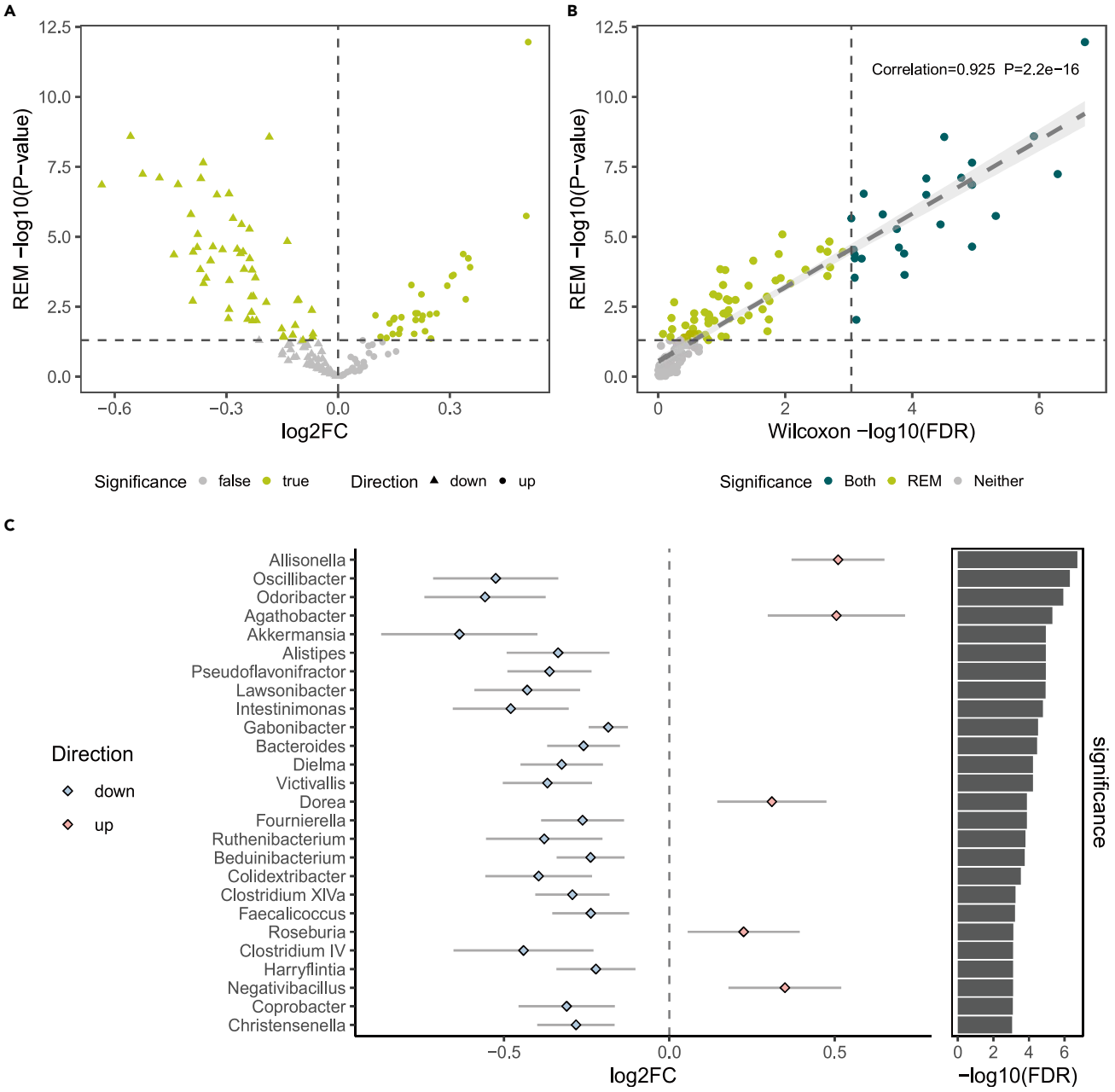


Figure 2. The differential abundance of microbial genera across all studies

(A) The volcano plot of the random effects model. REM: Random Effect Model. The horizontal dashed line represents the significant level of $p < 0.05$ of the REM. The vertical dashed line denotes the cutoff of the enrichment and depletion of genera. The triangles indicate the decrease of genera in obese patients whereas the circles indicate the increase of genera in obese patients. The significance of genera is represented by the light green or gray colors.

(B) The significant level of the REM and the Stratified Wilcoxon test. The correlation was calculated by the Spearman correlation test. The horizontal dashed line is the cutoff of the significant level of the REM with $p< 0.05$. The vertical dashed line is the cutoff of the significant level of the Stratified Wilcoxon with FDR <0.001. The light green points represent the genera that are only significant in the REM. The dark green points indicate that the genera are statistically significant according to both the REM and Stratified Wilcoxon FDR. The gray points represent genera that are not significant according to the REM and Stratified Wilcoxon FDR.

(C) The forest plot of the significant genera with REM $p < 0.05$ and Wilcoxon FDR < 0.001. The bar plot on the left indicates the significant level of Wilcoxon FDR. The gray line denotes the confidence interval.

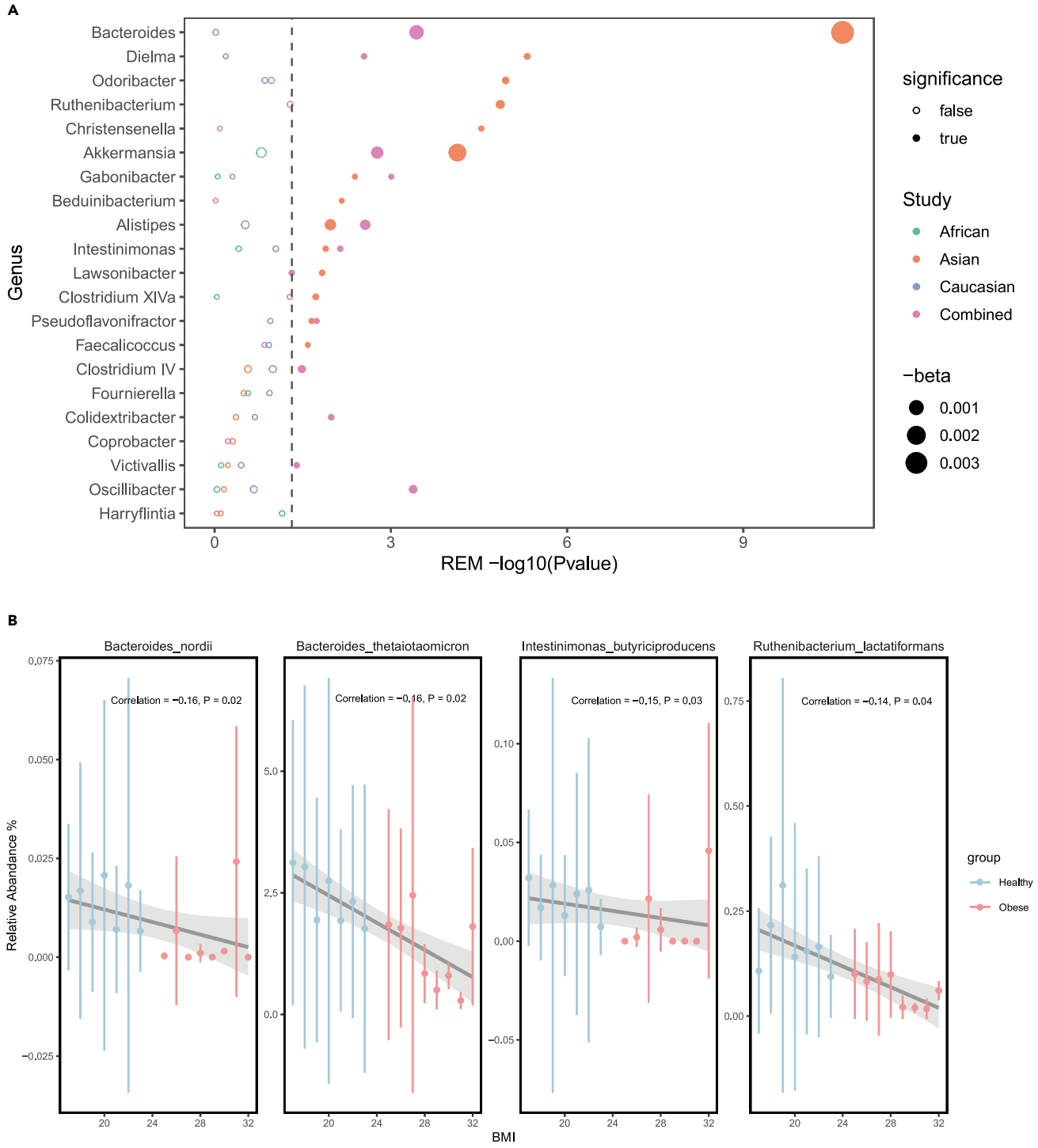


Figure 3. The correlation between genera and species with BMI

(A) The bubble chart of the correlation between the obesity-associated depleted genera and BMI. The beta coefficients were identified by the random effects model. The value on the x-axis represents the -$\log_{10}$ p-value from the random effects model. The significance of the beta coefficient is denoted by filled or unfilled bubbles. The effect size of the beta coefficients is indicated by the size of the bubbles.

(B) The correlation between significant species and BMI was identified by the Spearman correlation test using the shotgun metagenomic sequencing data from HKGutMicMap. The points with different colors indicate the obese or non-obese individuals. The lines represent the standard deviation of species abundance. The trend line was fitted by the "lm" method using the ggplot2 package.

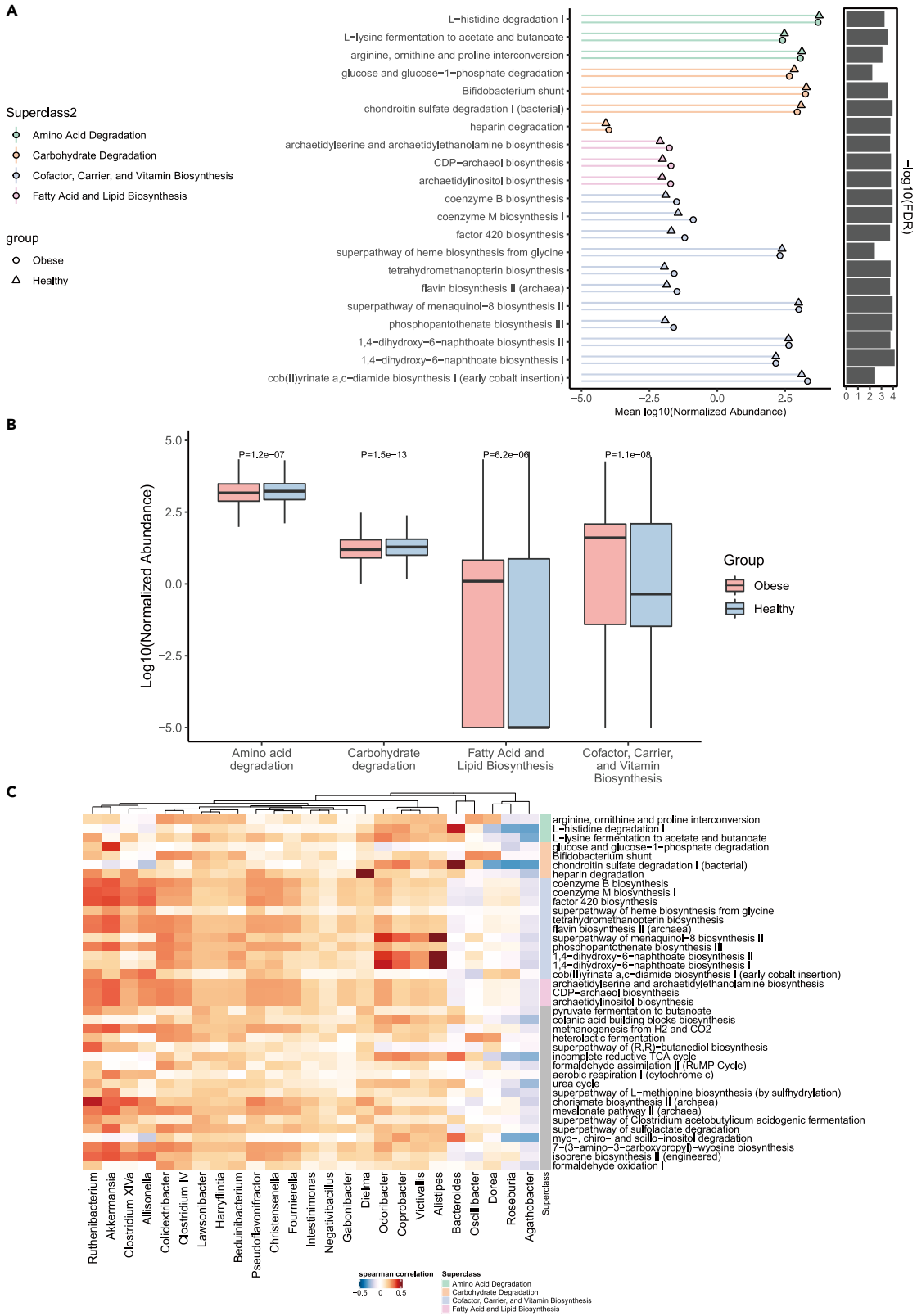

A
Superclass2
Amino Acid Degradation
Carbohydrate Degradation
Cofactor, Carrier, and Vitamin Biosynthesis
Fatty Acid and Lipid Biosynthesis
group
Obese
Healthy
L−histidine degradation I
L−lysine fermentation to acetate and butanoate
arginine, ornithine and proline interconversion
glucose and glucose−1−phosphate degradation
Bifidobacterium shunt
chondroitin sulfate degradation I (bacterial)
heparin degradation
archaetidylserine and archaetidylethanolamine biosynthesis
CDP−archaeol biosynthesis
archaetidylinositol biosynthesis
coenzyme B biosynthesis
coenzyme M biosynthesis I
factor 420 biosynthesis
superpathway of heme biosynthesis from glycine
tetrahydromethanopterin biosynthesis
flavin biosynthesis II (archaea)
superpathway of menaquinol−8 biosynthesis II
phosphopantothenate biosynthesis III
1,4−dihydroxy−6−naphthoate biosynthesis II
1,4−dihydroxy−6−naphthoate biosynthesis I
cob(II)yrinate a,c−diamide biosynthesis I (early cobalt insertion)
−5.0
−2.5
0.0
2.5
Mean log10(Normalized Abundance)
−log10(FDR)
0 1 2 3 4
B
Log10(Normalized Abundance)
5.0
2.5
0.0
−2.5
−5.0
P=1.2e−07
P=1.5e−13
P=6.2e−06
P=1.1e−08
Amino acid degradation
Carbohydrate degradation
Fatty Acid and Lipid Biosynthesis
Cofactor, Carrier, and Vitamin Biosynthesis
Group
Obese
Healthy
C
arginine, ornithine and proline interconversion
L−histidine degradation I
L−lysine fermentation to acetate and butanoate
glucose and glucose−1−phosphate degradation
Bifidobacterium shunt
chondroitin sulfate degradation I (bacterial)
heparin degradation
coenzyme B biosynthesis
coenzyme M biosynthesis I
factor 420 biosynthesis
superpathway of heme biosynthesis from glycine
tetrahydromethanopterin biosynthesis
flavin biosynthesis II (archaea)
superpathway of menaquinol−8 biosynthesis II
phosphopantothenate biosynthesis III
1,4−dihydroxy−6−naphthoate biosynthesis II
1,4−dihydroxy−6−naphthoate biosynthesis I
cob(II)yrinate a,c−diamide biosynthesis I (early cobalt insertion)
archaetidylserine and archaetidylethanolamine biosynthesis
CDP−archaeol biosynthesis
archaetidylinositol biosynthesis
pyruvate fermentation to butanoate
colanic acid building blocks biosynthesis
methanogenesis from H2 and CO2
heterolactic fermentation
superpathway of (R,R)−butanediol biosynthesis
incomplete reductive TCA cycle
formaldehyde assimilation II (RuMP Cycle)
aerobic respiration I (cytochrome c)
urea cycle
superpathway of L−methionine biosynthesis (by sulfhydrylation)
chorismate biosynthesis II (archaea)
mevalonate pathway II (archaea)
superpathway of Clostridium acetobutylicum acidogenic fermentation
superpathway of sulfolactate degradation
myo−, chiro− and scillo−inositol degradation
7−(3−amino−3−carboxypropyl)−wyosine biosynthesis
isoprene biosynthesis II (engineered)
formaldehyde oxidation I
Superclass
Ruthenibacterium
Akkermansia
Clostridium XlVa
Allisonella
Colidextribacter
Clostridium IV
Lawsonibacter
Harryflintia
Beduinibacterium
Pseudoflavonifractor
Christensenella
Fournierella
Intestinimonas
Negativibacillus
Gabonibacter
Dielma
Odoribacter
Coprobacter
Victivallis
Alistipes
Bacteroides
Oscillibacter
Dorea
Roseburia
Agathobacter
spearman correlation
−0.5 0 0.5
Superclass
Amino Acid Degradation
Carbohydrate Degradation
Cofactor, Carrier, and Vitamin Biosynthesis
Fatty Acid and Lipid Biosynthesis

Figure 4. The functional analysis between the obese and non-obese groups

(A) The lollipop plot shows the significant functional modules that were identified by the random effect model with $p < 0.05$ and Stratified Wilcoxon FDR < 0.01. The bar plot on the left indicates the significant level of Wilcoxon FDR. These modules could be detected in more than 9 studies. The triangular and circular points represent the mean $\log_{10}$ normalized abundance of functional pathways in obese and healthy groups.

(B) The four Superclass levels of functional modules were compared between the obese and non-obese groups. The significance between obesity groups was identified by the Wilcoxon test based on the raw abundance of each functional pathway.

(C) The association between the significant genera and the significant pathways. The bar on the right shows the categories of these pathways.

belonging to obesity-associated enriched genera, none of them were positively correlated with BMI (Figure S11B and Table S13).

Association of predicted microbiome functions with obesity

From 16S gut microbiota compositions, we inferred functional pathways using PICRUSt2 and applied random effects models and Stratified Wilcoxon tests to identify microbiome functions associated with obesity. 21 functional modules belonging to four categories including amino acid degradation, carbohydrate degradation, cofactor, carrier and vitamin biosynthesis, fatty acid, and lipid biosynthesis were significantly different between obese and lean individuals ($p<0.05$ random effects model & FDR <0.01 Stratified Wilcoxon test) (Figure 4A and Table S14). Among all functional pathways with significant differences between obese and healthy groups (Table S14), two fermentation-related pathways – the superpathway of *Clostridium acetobutylicum* acidogenic fermentation ($\log_{10}$ abundance in obese: 3.099 versus healthy: 3.181) and pyruvate fermentation to butanoate ($\log_{10}$ abundance in obese: 3.008 versus healthy: 3.092) – were decreased in obese subjects compared to healthy controls. These two pathways are capable of producing the short chain fatty acids (SCFAs) acetate and butyrate by fermenting a variety of fibers and sugars. The function alteration in obesity suggested the gut microbiome in obese subjects produced fewer SCFAs. Although SCFAs might increase intestinal energy harvesting, they could also increase energy expenditure, regulate appetite, increase insulin sensitivity and reduce bodyweight.[43] Obese individuals had lower levels of amino acid and carbohydrate degradation and higher levels of fatty acid and lipid biosynthesis and cofactor, carrier and vitamin biosynthesis pathways relative to healthy individuals (Figure 4B). This result is in accordance with the hypothesis that obese individuals tend to have high-fat diets with low carbohydrates and proteins, which reflects the reduced carbohydrate and protein utilization and increased lipid synthesis functions in the gut microbiome. In addition, we calculated correlations between functional pathways and microbial genera and found three obesity-associated enriched genera (*Dorea*, *Roseburia*, and *Agathobacter*) were negatively correlated with several functional modules closely related to obesity, such as chondroitin sulfate degradation and L-histidine degradation (Figure 4C). The products of these two modules have been reported to be lower in the obese human gut compared with non-obese individuals.[44,45] Taken together, these observations indicate that functional changes in obese individuals might be because of obesity-related diets that induce compositional changes in gut microbial communities.

Machine learning model to classify obese and healthy individuals

In addition to the statistical methods described above (random effects model and Stratified Wilcoxon test), we assessed the performances of machine learning models in categorizing obese and non-obese individuals based on differentially abundant microbial taxa. To identify the optimal model for obesity classification based on all microbial genera, we applied three frequently used machine learning methods, XGBoost, neural network, and random forest, to construct the models. Compared to the neural network and random forest models, XGBoost has the best AUC (Table 2 and Figures S12A–S12C), suggesting that XGBoost has a greater generalization ability in predicting obesity using the microbial features from different studies. Therefore, we applied the XGBoost model for further analysis. As our study cohorts consisted of populations from various backgrounds, there might be a covariate shift issue in the dataset. The covariate shift represents a situation when the distribution of training data is highly different from testing data. The models cannot give a satisfactory predictive performance if the covariate shift is presented in the dataset.[46] Therefore, we corrected the covariate shift in the dataset and trained the XGBoost models. From the models based on all cohorts (Table 3 and Figures S13A–S13C), we found that the models with covariate shift correction had better discrimination than models without covariate shift correction (median AUC (Area Under the ROC Curve): 0.608 versus 0.605, 10-fold cross-validation; median AUC: 0.625 versus 0.606, leave-one-study-out validation). However, the relatively low AUC indicated that the models did not have sufficient discriminatory power to classify obese and healthy groups, which was consistent with Sze et al.,[21] who reported that machine learning models based on microbial taxa were insufficient to

Table 2. The comparison of model performance measured by median AUC(IQR) based on XGBoost, neural network, and randomforest

Training strategies	XGBoost	Neural network	Randomforest
Cross validation	0.605(0.015)	0.588(0.018)	0.550(0.012)
Within study	0.640(0.180)	0.578(0.168)	0.585(0.169)
Leave-one-study-out validation	0.606(0.087)	0.575(0.075)	0.520(0.043)

AUC: Area under the ROC Curve.
IQR: Interquartile Range.

discriminate between obese and healthy individuals. Within studies, we found that those designed specifically for investigating obesity and gut microbiota associations had better model discrimination compared to gut microbiota survey studies of general populations (Figure S13C). We therefore refined our model to only include obesity association studies and applied the same model training strategies. The median AUC of the XGBoost model trained in the dataset corrected covariate shift was 0.771 using 10-fold cross-validation (Table 4 and Figures S14A–S14C). Correcting covariate shift in the testing dataset could improve the model performance to some extent. However, intrinsic data characteristics still have a large influence on the final results of model classification, even though covariate shift was accounted for in the model training.

The permutation importance of each microbial feature was identified based on the XGBoost model trained on the data of obesity association studies using 10-fold cross-validation. Among the microbial genera with the top 30 permutation importance (Figure S15), 10 microbial genera (*Akkermansia, Alistipes, Agathobacter, Allisonella, Bacteroides, Dorea, Oscillibacter, Clostridium IV, Negativibacillus, and Ruthenibacterium*) were overlapped with the 26 core taxa selected from random effects model and Stratified Wilcoxon test. Among these important genera, *Akkermansia* had the lowest $\log_2$ fold change compared to other genera, indicating this genus was highly deficient in the obese human gut. *Blautia*, which was more abundant in obese subjects, had the highest permutation importance in the XGBoost model. The role of *Blautia* in obesity is controversial as some studies have reported a positive association between *Blautia* and visceral fat,[47,48] whereas other studies demonstrated that some species in the *Blautia* genus are depleted in obesity.[49]

DISCUSSION

There is limited evidence demonstrating how changes in the composition of the gut microbiome contribute to obesity. In this study, we combined fecal microbiome community data from 18 studies consisting of more than 7,000 individuals to evaluate relationships between the gut microbiome and obesity. We found (1) a decreased alpha diversity in the obesity gut microbiome compared with healthy lean controls across 18 studies, (2) several bacterial genera implicated in obesity that were depleted in obese individuals and more abundant in healthy controls, and (3) novel evidence of obesity-related functional changes, including enriched lipid biosynthesis and depleted carbohydrate and protein degradation pathways in obese individuals. Taken together, these results indicate that obesity is associated with reduced gut microbial community diversity, likely driven by a lack of commensal microbes and characterized by high fat, low-carbohydrate, and low-protein metabolism. These findings reveal consistent microbial signatures for obesity across 18 independent cohorts encompassing various host backgrounds.

Table 3. The XGBoost model performance measured by median AUC(IQR) with and without covariate shift correction on the overall population

Training strategies	Without covariate shift correction	With covariate shift correction
Cross validation	0.605(0.015)	0.608(0.018)
Within study	0.640(0.180)	0.636(0.166)
Leave-one-study-out validation	0.606(0.087)	0.625(0.106)

AUC: Area under the ROC Curve.
IQR: Interquartile Range.

Table 4. The XGBoost model performance measured by median AUC(IQR) with and without covariate shift correction on the population from obesity association studies

Training strategies	Without covariate shift correction	With covariate shift correction
Cross validation	0.768(0.058)	0.771(0.075)
Within study	0.805(0.128)	0.862(0.179)
Leave-one-study-out validation	0.667(0.073)	0.689(0.093)

AUC: Area under the ROC Curve.
IQR: Interquartile Range.

Our analysis revealed a decreased alpha diversity in obese individuals relative to healthy controls, which is consistent with a previous meta-analysis published by Sze et al.[21] despite the larger number of cohorts, sample size and ethnic background diversity in this present study. This result further supports the notion that there is distinct shift in the gut microbiome composition in obese individuals compared with healthy individuals. Le Chatelier et al.[50] investigated the microbiome composition in obese and healthy Danes and found that individuals with a lower gut microbiota richness were more likely to be obese and have insulin resistance, dyslipidemia, and higher inflammatory levels compared with individuals with higher gut microbiota richness. These consistent patterns of an obesity gut microbiome open the possibility of managing obesity via modifying the gut microbiome, either by prebiotic or probiotic interventions that enrich beneficial bacteria in the gut.

A crucial finding of this study showed that a typical feature of gut microbiota in obese individuals was the depletion of several commensal microbial genera. Of the microbial genera negatively associated with BMI, *Bacteroides*, *Akkermansia*, *Alistipes*, *Odoribacter*, *Oscillibacter*, and *Intestinimonas* have been previously reported to be reduced in obese individuals in several studies.[18,51–53] Among these genera, *Bacteroides* and *Alistipes* are prevalent in the healthy human gut and are considered potential opportunistic pathogens in diseases such as bacteremia, appendicitis and rectal abscesses.[54,55] However, there are opposing views that they may have protective effects against diseases such as cardiovascular disease and colitis.[56,57] *Odoribacter* and *Oscillibacter*, together with *Akkermansia*, may have beneficial effects in metabolic disorders and ulcerative colitis.[58–60] *Odoribacter splanchnicus* is considered a probiotic species that produces short chain fatty acids including butyrate.[61] For the *Intestinimonas* genus, we identified *I. butyriciproducens* in the HKGutMicMap cohort to be negatively associated with BMI. *I. butyriciproducens*, also a butyrate producer, is a beneficial bacterium that was reduced in aged mice with insulin resistance relative to control mice.[62] Another species negatively correlated with BMI, *R. lactatiformans*, was reported to be able to reduce inflammation and stimulate immune function.[63] Other BMI associated genera and species identified in our study were less often reported in the literature and thus warrant further investigation. We noted that among the differentially abundant microbial genera in obese and healthy individuals, most had a negative association with BMI in the combined 18 populations examined and this trend was more apparent in the Asian population compared with Caucasians and Africans. One explanation is that distinct ethnic backgrounds, diets, habits, socioeconomic status, and urbanization could impact how the gut microbiome is associated with obesity across populations.[64] Our subgroup analysis in different age groups indicated that most taxa were significantly associated with BMI in adults, but no association was observed in children. The relatively small sample sizes in the child population may have limited the inference of associations between microbial taxa and BMI. In populations stratified by sex, we found inconsistent associations of taxa and BMI in males and females. As not all studies recorded the sex information (males were found in 12 studies and females were found in 11 studies), the possibility of inconsistent associations may be because of interstudy variations that confound taxa abundance and BMI associations in different sex groups.

We identified several pathways, including amino acid degradation, carbohydrate degradation, and fatty acid and lipid biosynthesis related to obesity. Among the amino acid degradation pathways, L-histidine degradation I was depleted in obese individuals. A systematic review proposed that histidine supplementation is able to ameliorate metabolic syndrome and inhibit food intake, according to previous human and animal studies.[44] In addition to amino acid degradation, chondroitin sulfate degradation (a carbohydrate degradation pathway) was also lower in obese subjects relative to healthy controls. Clinical and rodent studies have reported that chondroitin sulfate has an anti-inflammatory effect and prevents the formation of atheroma in obese individuals.[45,65] In the Superclass level of functional pathways, the obese group showed decreased levels of amino acid degradation, carbohydrate degradation, and increased levels of

fatty acid and lipid biosynthesis. Changes in these functional pathways was attributable to the differentially abundant microbial taxa between obese individuals and healthy controls, indicating that the obese gut microbiome adapts to a diet with a higher proportion of fat and less amino acids and carbohydrates. There are two major classes of metabolites in the colon that are derived from microorganisms utilizing dietary substrates: secondary bile acids and SCFAs.[66] One possible result of westernized high-fat and low-fiber diets is that the gut microbes cannot produce sufficient SCFAs from fiber fermentation as an energy source and shift to utilize dietary fat, which increases the products from microbial fatty acid and lipid biosynthesis such as bile acids and secondary bile acids.[67] With a pro-inflammatory effect, the secondary bile acids are considered as risk factors for liver inflammation and colorectal cancer development.[68,69] Obesity has been described as a systemic chronic low-grade inflammation in response to overnutrition and excessive energy.[70] Higher dietary fat may lead to the increased biosynthesis of unfavorable lipid metabolites that contribute to chronic intestinal inflammation status in obesity. Growing evidence suggests that high-fat diets are associated with increased levels of inflammatory cytokines.[71] Among the obesity-associated depleted genera identified here, *Oscillospira*,[72] *Odoribacter*,[61] *Akkermansia*,[73] and *Ruthenibacterium*[63] are known anti-inflammatory taxa with the capability of reducing intestinal inflammation. Carrio et al.[74] reported that *Akkermansia* was negatively associated with pro-inflammatory cytokines such as IL-6 in the serum. In contrast, the genus *Allisonella* enriched in obesity promotes inflammation by producing higher levels of histamine and other inflammatory molecules.[75] Based on existing evidence, we posit that the obese gut microbiome composition might be influenced by high-fat diets and is characterized by lower and higher relative abundances of anti-inflammatory and pro-inflammatory bacteria, respectively.

Compared to models trained on an overall population consisting of microbiome surveys and obesity association studies, we found that models trained on populations from obesity association studies had improved model discrimination, suggesting that classification models based on generic population surveys do not have the sensitivity to detect obesity-related patterns. Compared with meta-analysis published by Sze et al.,[21] our classification model based on large scale populations from obesity association studies showed better classification performance (AUC of leave-one-study-out validation: 0.689 versus 0.567). However, the performance is still insufficient compared to those of classification models for other diseases such as CRC or inflammatory bowel disease (IBD).[76] A previous CRC meta-analysis[77] identified 29 microbial taxa with Stratified Wilcoxon FDR < 1e-5, and AUCs for leave-one-study-out validation ranged between 0.71 and 0.91. In our study, the core taxa we included in the meta-analysis were 26 genera with Stratified Wilcoxon FDR < 0.001; only four of them (*Allisonella*, *Oscillibacter*, *Odoribacter*, and *Agathobacter*) had an FDR < 1e-5. The lower FDR value implicated the higher discriminatory power of microbial features in CRC than in obesity for disease classification. As a complex multifactorial disease, obesity is a product of the interplay of genetics, epigenetics, environment and microbiome, with a high possibility of coexisting with metabolic disorders such as diabetes, hypertension, and cardiovascular diseases.[78] The inherently complicated pathophysiology of obesity makes it difficult to achieve a high AUC based on microbial features compared to other gastrointestinal diseases. In addition, based on populations from 18 studies consisting of diverse ethnic backgrounds, our study may have larger study heterogeneity than the CRC meta-analysis with eight studies included. The high variation of obesity datasets may limit the performance of the classification model resulting in a relatively low AUC. However, the model discrimination improved compared to the model from prior obesity meta-analysis, especially in the subset of studies designed for investigating obesity-microbiome association. Although our study indicated generalizable microbial markers for obesity, all these microbial markers should be validated in well-designed cohorts across various populations.

Compared with published meta-analyses on the relationship between obesity and microbiome, one strength of our study is that we included data from Asian populations (He et al., HKGutMicMap) resulting in a more diverse ethnic backgrounds compared to a previous meta-analysis by Stanislawski et al.[22] We further evaluated the relationship between microbial taxa and BMI and identified a novel set of genera associated with BMI, which might be signatures for obesity-related microbiome disorders. In addition, we trained a better obesity classification model based on selected significant genera compared the study by Sze et al.,[21] suggesting that the selected taxa contributed to distinct gut microbiome composition of obesity. Our study is an important extension of previous meta-analysis for obesity and reveals potential mechanisms that link gut microbiome composition to obesity-related conditions through the microbial functional analysis.

In obese human guts, we found decreased alpha diversity, a lack of commensal bacteria, and specific metabolic patterns, indicating potential microbiota dysbiosis. Compared to obesity classification models trained

on the overall population from the 18 studies, we found models trained on populations from obesity-microbiome association studies have improved discrimination. Notably, the commensal species depleted in obesity (*B. nordii*, *B. thetaiotaomicron*, *I. butyriciproducens*, and *R. lactatiformans*) identified in this meta-analysis have the potential to be considered in developing novel probiotics targeted to obesity. As our understanding of the obesity gut microbiome increases, determining and modulating modifiable features may help reduce the rising trend of obesity, as well as exert a positive influence on other metabolic diseases.

Limitations of the study

There are several limitations in this study. As we combined different public datasets to identify relationships between obesity and microbiome, the comparison between studies is likely constrained by geographic and technical variation. In addition, the included studies applied 16S amplicon sequencing strategies which do not reliably provide species and strain level taxonomic resolution. Similarly, functional modules predicted by the PICRUSt2 cannot distinguish species- or strain-level functionality. To overcome some of these limitations, we applied the random effect model to adjust the study variation and use shotgun metagenomic sequencing data from the HKGutMicMap study to identify putative species level markers according to the significant genera identified in meta-analysis. The associations reported here identify several bacterial genera and species that likely play roles in obesity development. The next step is to investigate any causal effects these taxa may have on obesity using well-phenotyped longitudinal cohorts and animal models to reveal potential host-microbe interactions and mechanisms.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Literature search
  - Bacterial community diversity analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Differential abundance analysis of taxa and functional modules
  - Species identification using shotgun metagenome data
  - Machine learning models for obesity classification

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.106476.

ACKNOWLEDGMENTS

This study was funded by InnoHK, The Government of Hong Kong, Special Administrative Region of the People's Republic of China. The funder of the study had no role in study design, data collection, data analysis, data interpretation, or manuscript writing. Data from the HMP[38] and Zupancic et al.[30] study were obtained from the National Institutes of Health (NIH)-designated data repositories, the database of Genotypes and Phenotypes (dbGaP), under the accession numbers of phs000228.v4.p1 and phs000258.v2.p1, respectively. For the HMP[38] dataset included in this study, funding support for the development of NIH Human Microbiome Project - Core Microbiome Sampling Protocol A (HMP-A) was provided by the NIH Roadmap for Medical Research. Clinical data from HMP study were jointly produced by the Baylor College of Medicine and the Washington University School of Medicine. Sequencing data were produced by the Baylor College of Medicine Human Genome Sequencing Center, The Broad Institute, the Genome Center at Washington University, and the J. Craig Ventor Institute. These data were submitted by the EMMES Corporation, which serves as the clinical data collection site for the HMP. For the Zupancic et al.[30] dataset included in this study, it was supported by UH2/UH3 award DK83982 from the National Institutes of Health to CMF-L and ARS, and U01 GM074518 and P30 DK072488 (Mid-Atlantic Nutrition and Obesity Research Center) to ARS. Data from the studies of Ahmad et al.,[23] Chavez et al.,[24] Barengolts et al.,[25] Gao et al.,[26] Olsson et al.,[27] Lippert et al.,[28] Ross et al.,[29] Escobar et al.,[31] Wu et al.,[32] Zeevi et al.,[33] Goodrich et al.,[34]

Baxter et al.,[35] HKGutMicMap,[36] He et al.,[37] AGP,[39] and Stanislawski et al.[40] were obtained from the Sequence Read Archive (SRA) or Europe Nucleotide Archive (ENA) database. We thank Matthew Wong for his suggestions to improve the language in the manuscript.

AUTHOR CONTRIBUTIONS

Study design, Y.L., Z.X., Y.K.Y., and S.C.N.; Literature review, Y.L., Z.X., W.J., and W.H.; Methodology, Y.L., Z.X., Y.K.Y., and H.M.T.; Data analysis, Y.L.; Writing – Original draft, Y.L.; Writing – Review and editing: Z.X., Y.K.Y., H.M.T., and S.C.N.; Funding Acquisition: S.C.N.; Supervision, S.C.N. and F.K.L.C.

DECLARATION OF INTERESTS

F.K.L.C. is the co-founder, non-executive Board Chairman and shareholder of GenieBiome Ltd. F.K.L.C. is Board Member of CUHK Medical Center. F.K.L.C. has received fees as an advisor and honoraria as a speaker for Eisai Co. Ltd., AstraZeneca, Pfizer Inc., Takeda Pharmaceutical Co., and Takeda (China) Holdings Co. Ltd. F.K.L.C. receives patent royalties through his affiliated institutions in the applications of microbiome. S.C.N. is a scientific co-founder and shareholder of GenieBiome Ltd. S.C.N. has served as an advisory board member for Pfizer, Ferring, Janssen, and Abbvie and received honoraria as a speaker for Ferring, Tillotts, Menarini, Janssen, Abbvie, and Takeda. S.C.N. has received research grants through her affiliated institutions from Olympus, Ferring, and Abbvie. S.C.N. receives patent royalties through her affiliated institutions in the applications of microbiome. Z.X. is part-time employee of GenieBiome Ltd. S.C.N., F.K.L.C., and Z.X. are named inventors of patent applications held by the CUHK and MagIC that covers the therapeutic and diagnostic use of microbiome related to obesity. All other co-authors declare no competing interests.

Received: October 1, 2022
Revised: January 14, 2023
Accepted: March 20, 2023
Published: March 22, 2023

REFERENCES

1. Flegal, K.M., Kit, B.K., Orpana, H., and Graubard, B.I. (2013). Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis. JAMA *309*, 71–82. https://doi.org/10.1001/jama.2012.113905.

2. Sivamaruthi, B.S., Kesika, P., Suganthy, N., and Chaiyasut, C. (2019). A review on role of microbiome in obesity and antiobesity properties of probiotic supplements. BioMed Res. Int. *2019*, 3291367. https://doi.org/10.1155/2019/3291367.

3. Bouter, K.E., van Raalte, D.H., Groen, A.K., and Nieuwdorp, M. (2017). Role of the gut microbiome in the pathogenesis of obesity and obesity-related metabolic dysfunction. Gastroenterology *152*, 1671–1678. https://doi.org/10.1053/j.gastro.2016.12.048.

4. Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., et al. (2009). A core gut microbiome in obese and lean twins. Nature *457*, 480–484. https://doi.org/10.1038/nature07540.

5. Ridaura, V.K., Faith, J.J., Rey, F.E., Cheng, J., Duncan, A.E., Kau, A.L., Griffin, N.W., Lombard, V., Henrissat, B., Bain, J.R., et al. (2013). Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science *341*, 1241214. https://doi.org/10.1126/science.1241214.

6. Vrieze, A., Van Nood, E., Holleman, F., Salojärvi, J., Kootte, R.S., Bartelsman, J.F.W.M., Dallinga-Thie, G.M., Ackermans, M.T., Serlie, M.J., Oozeer, R., et al. (2012). Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. Gastroenterology *143*, 913–916.e7. https://doi.org/10.1053/j.gastro.2012.06.031.

7. Ng, S.C., Xu, Z., Mak, J.W.Y., Yang, K., Liu, Q., Zuo, T., Tang, W., Lau, L., Lui, R.N., Wong, S.H., et al. (2022). Microbiota engraftment after faecal microbiota transplantation in obese subjects with type 2 diabetes: a 24-week, double-blind, randomised controlled trial. Gut *71*, 716–723. https://doi.org/10.1136/gutjnl-2020-323617.

8. Mocanu, V., Zhang, Z., Deehan, E.C., Kao, D.H., Hotte, N., Karmali, S., Birch, D.W., Samarasinghe, K.K., Walter, J., and Madsen, K.L. (2021). Fecal microbial transplantation and fiber supplementation in patients with severe obesity and metabolic syndrome: a randomized double-blind, placebo-controlled phase 2 trial. Nat. Med. *27*, 1272–1279. https://doi.org/10.1038/s41591-021-01399-2.

9. Sun, L., Ma, L., Ma, Y., Zhang, F., Zhao, C., and Nie, Y. (2018). Insights into the role of gut microbiota in obesity: pathogenesis, mechanisms, and therapeutic perspectives. Protein Cell *9*, 397–403. https://doi.org/10.1007/s13238-018-0546-3.

10. Goffredo, M., Mass, K., Parks, E.J., Wagner, D.A., McClure, E.A., Graf, J., Savoye, M., Pierpont, B., Cline, G., and Santoro, N. (2016). Role of gut microbiota and short chain fatty acids in modulating energy harvest and fat partitioning in youth. J. Clin. Endocrinol. Metab. *101*, 4367–4376. https://doi.org/10.1210/jc.2016-1797.

11. Tims, S., Derom, C., Jonkers, D.M., Vlietinck, R., Saris, W.H., Kleerebezem, M., de Vos, W.M., and Zoetendal, E.G. (2013). Microbiota conservation and bmi signatures in adult monozygotic twins. ISME J. *7*, 707–717. https://doi.org/10.1038/ismej.2012.146.

12. Jumpertz, R., Le, D.S., Turnbaugh, P.J., Trinidad, C., Bogardus, C., Gordon, J.I., and Krakoff, J. (2011). Energy-balance studies reveal associations between gut microbes, caloric load, and nutrient absorption in humans. Am. J. Clin. Nutr. *94*, 58–65. https://doi.org/10.3945/ajcn.110.010132.

13. Yasir, M., Angelakis, E., Bibi, F., Azhar, E.I., Bachar, D., Lagier, J.C., Gaborit, B., Hassan, A.M., Jiman-Fatani, A.A., Alshali, K.Z., et al. (2015). Comparison of the gut microbiota of people in France and Saudi Arabia. Nutr. Diabetes *5*, e153. https://doi.org/10.1038/nutd.2015.3.

14. Nirmalkar, K., Murugesan, S., Pizano-Zarate, M.L., Villalobos-Flores, L.E., Garcia-Gonzalez, C., Morales-Hernandez, R.M., Nunez-Hernandez, J.A., Hernandez-Quiroz,

F., Romero-Figueroa, M.D.S., Hernandez-Guerrero, C., et al. (2018). Gut microbiota and endothelial dysfunction markers in obese mexican children and adolescents. Nutrients *10*, 2009. https://doi.org/10.3390/nu10122009.

15. Gao, R., Zhu, C., Li, H., Yin, M., Pan, C., Huang, L., Kong, C., Wang, X., Zhang, Y., Qu, S., et al. (2018). Dysbiosis signatures of gut microbiota along the sequence from healthy, young patients to those with overweight and obesity. Obesity *26*, 351–361. https://doi.org/10.1002/oby.22088.

16. Ville, A., Levine, E., Zhi, D., Lararia, B., and Wojcicki, J.M. (2020). Alterations in the gut microbiome at 6 months of age in obese latino infants. J. Am. Coll. Nutr. *39*, 47–53. https://doi.org/10.1080/07315724.2019.1606744.

17. Bai, J., Hu, Y., and Bruner, D.W. (2019). Composition of gut microbiota and its association with body mass index and lifestyle factors in a cohort of 7-18 years old children from the american gut project. Pediatr. Obes. *14*, e12480. https://doi.org/10.1111/ijpo.12480.

18. Hu, H.J., Park, S.G., Jang, H.B., Choi, M.K., Park, K.H., Kang, J.H., Park, S.I., Lee, H.J., and Cho, S.H. (2015). Obesity alters the microbial community profile in Korean adolescents. PLoS One *10*, e0134333. https://doi.org/10.1371/journal.pone.0134333.

19. Chen, X., Sun, H., Jiang, F., Shen, Y., Li, X., Hu, X., Shen, X., and Wei, P. (2020). Alteration of the gut microbiota associated with childhood obesity by 16s rrna gene sequencing. PeerJ *8*, e8317. https://doi.org/10.7717/peerj.8317.

20. Dai, Z., Coker, O.O., Nakatsu, G., Wu, W.K.K., Zhao, L., Chen, Z., Chan, F.K.L., Kristiansen, K., Sung, J.J.Y., Wong, S.H., et al. (2018). Multi-cohort analysis of colorectal cancer metagenome identified altered bacteria across populations and universal bacterial markers. Microbiome *6*, 70. https://doi.org/10.1186/s40168-018-0451-2.

21. Sze, M.A., and Schloss, P.D. (2016). Looking for a signal in the noise: revisiting obesity and the microbiome. mBio *7*, 01018-16. https://doi.org/10.1128/mBio.01018-16.

22. Stanislawski, M.A., Dabelea, D., Lange, L.A., Wagner, B.D., and Lozupone, C.A. (2019). Gut microbiota phenotypes of obesity. NPJ Biofilms Microbiomes *5*, 18. https://doi.org/10.1038/s41522-019-0091-8.

23. Ahmad, A., Yang, W., Chen, G., Shafiq, M., Javed, S., Ali Zaidi, S.S., Shahid, R., Liu, C., and Bokhari, H. (2019). Analysis of gut microbiota of obese individuals with type 2 diabetes and healthy individuals. PLoS One *14*, e0226372. https://doi.org/10.1371/journal.pone.0226372.

24. Chávez-Carbajal, A., Nirmalkar, K., Pérez-Lizaur, A., Hernández-Quiroz, F., Ramírez-Del-Alto, S., García-Mena, J., and Hernández-Guerrero, C. (2019). Gut microbiota and predicted metabolic pathways in a sample of mexican women affected by obesity and obesity plus metabolic syndrome. Int. J. Mol. Sci. *20*, 438. https://doi.org/10.3390/ijms20020438.

25. Barengolts, E., Green, S.J., Eisenberg, Y., Akbar, A., Reddivari, B., Layden, B.T., Dugas, L., and Chlipala, G. (2018). Gut microbiota varies by opioid use, circulating leptin and oxytocin in african american men with diabetes and high burden of chronic disease. PLoS One *13*, e0194171. https://doi.org/10.1371/journal.pone.0194171.

26. Gao, X., Jia, R., Xie, L., Kuang, L., Feng, L., and Wan, C. (2018). A study of the correlation between obesity and intestinal flora in school-age children. Sci. Rep. *8*, 14511. https://doi.org/10.1038/s41598-018-32730-6.

27. Olsson, L.M., Poitou, C., Tremaroli, V., Coupaye, M., Aron-Wisnewsky, J., Bäckhed, F., Clément, K., and Caesar, R. (2020). Gut microbiota of obese subjects with prader-willi syndrome is linked to metabolic health. Gut *69*, 1229–1238. https://doi.org/10.1136/gutjnl-2019-319322.

28. Lippert, K., Kedenko, L., Antonielli, L., Kedenko, I., Gemeier, C., Leitner, M., Kautzky-Willer, A., Paulweber, B., and Hackl, E. (2017). Gut microbiota dysbiosis associated with glucose metabolism disorders and the metabolic syndrome in older adults. Benef. Microbes *8*, 545–556. https://doi.org/10.3920/BM2016.0184.

29. Ross, M.C., Muzny, D.M., McCormick, J.B., Gibbs, R.A., Fisher-Hoch, S.P., and Petrosino, J.F. (2015). 16s gut community of the cameron county hispanic cohort. Microbiome *3*, 7. https://doi.org/10.1186/s40168-015-0072-y.

30. Zupancic, M.L., Cantarel, B.L., Liu, Z., Drabek, E.F., Ryan, K.A., Cirimotich, S., Jones, C., Knight, R., Walters, W.A., Knights, D., et al. (2012). Analysis of the gut microbiota in the old order amish and its relation to the metabolic syndrome. PLoS One *7*, e43052. https://doi.org/10.1371/journal.pone.0043052.

31. Escobar, J.S., Klotz, B., Valdes, B.E., and Agudelo, G.M. (2014). The gut microbiota of colombians differs from that of americans, europeans and asians. BMC Microbiol. *14*, 311. https://doi.org/10.1186/s12866-014-0311-6.

32. Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.Y., Keilbaugh, S.A., Bewtra, M., Knights, D., Walters, W.A., Knight, R., et al. (2011). Linking long-term dietary patterns with gut microbial enterotypes. Science *334*, 105–108. https://doi.org/10.1126/science.1208344.

33. Zeevi, D., Korem, T., Zmora, N., Israeli, D., Rothschild, D., Weinberger, A., Ben-Yacov, O., Lador, D., Avnit-Sagi, T., Lotan-Pompan, M., et al. (2015). Personalized nutrition by prediction of glycemic responses. Cell *163*, 1079–1094. https://doi.org/10.1016/j.cell.2015.11.001.

34. Goodrich, J.K., Waters, J.L., Poole, A.C., Sutter, J.L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J.T., et al. (2014). Human genetics shape the gut microbiome. Cell *159*, 789–799. https://doi.org/10.1016/j.cell.2014.09.053.

35. Baxter, N.T., Ruffin, M.T., 4th, Rogers, M.A.M., and Schloss, P.D. (2016). Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. Genome Med. *8*, 37. https://doi.org/10.1186/s13073-016-0290-3.

36. Yeoh, Y.K., Chen, Z., Wong, M.C.S., Hui, M., Yu, J., Ng, S.C., Sung, J.J.Y., Chan, F.K.L., and Chan, P.K.S. (2020). Southern Chinese populations harbour non-nucleatum fusobacteria possessing homologues of the colorectal cancer-associated fada virulence factor. Gut *69*, 1998–2007. https://doi.org/10.1136/gutjnl-2019-319635.

37. He, Y., Wu, W., Wu, S., Zheng, H.M., Li, P., Sheng, H.F., Chen, M.X., Chen, Z.H., Ji, G.Y., Zheng, Z.D.X., et al. (2018). Linking gut microbiota, metabolic syndrome and economic status based on a population-level analysis. Microbiome *6*, 172. https://doi.org/10.1186/s40168-018-0557-6.

38. Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207–214. https://doi.org/10.1038/nature11234.

39. McDonald, D., Hyde, E., Debelius, J.W., Morton, J.T., Gonzalez, A., Ackermann, G., Aksenov, A.A., Behsaz, B., Brennan, C., Chen, Y., et al. (2018). American gut: an open platform for citizen science microbiome research. mSystems *3*, 00031-18. https://doi.org/10.1128/mSystems.00031-18.

40. Stanislawski, M.A., Lozupone, C.A., Wagner, B.D., Eggesbø, M., Sontag, M.K., Nusbacher, N.M., Martinez, M., and Dabelea, D. (2018). Gut microbiota in adolescents and the association with fatty liver: the epoch study. Pediatr. Res. *84*, 219–227. https://doi.org/10.1038/pr.2018.32.

41. Sullivan, L.M., Dukes, K.A., and Losina, E. (1999). Tutorial in biostatistics - an introduction to hierarchical linear modelling. Stat. Med. *18*, 855–888. https://doi.org/10.1002/(Sici)1097-0258.

42. Gibbons, S.M., Duvallet, C., and Alm, E.J. (2018). Correcting for batch effects in case-control microbiome studies. PLoS Comput. Biol. *14*, e1006102. https://doi.org/10.1371/journal.pcbi.1006102.

43. Canfora, E.E., Jocken, J.W., and Blaak, E.E. (2015). Short-chain fatty acids in control of body weight and insulin sensitivity. Nat. Rev. Endocrinol. *11*, 577–591. https://doi.org/10.1038/nrendo.2015.128.

44. DiNicolantonio, J.J., McCarty, M.F., and OKeefe, J.H. (2018). Role of dietary histidine in the prevention of obesity and metabolic

syndrome. Open Heart *5*, e000676. https://doi.org/10.1136/openhrt-2017-000676.

45. de Abajo, F.J., Gil, M.J., García Poza, P., Bryant, V., Oliva, B., Timoner, J., and García-Rodríguez, L.A. (2014). Risk of nonfatal acute myocardial infarction associated with non-steroidal antiinflammatory drugs, non-narcotic analgesics and other drugs used in osteoarthritis: a nested case-control study. Pharmacoepidemiol. Drug Saf. *23*, 1128–1138. https://doi.org/10.1002/pds.3617.

46. Tripuraneni, N., Adlam, B., and Pennington, J. (2021). Covariate shift in high-dimensional random feature regression. Preprint at arXiv. https://doi.org/10.48550/arXiv.2111.08234.

47. Le Roy, C.I., Beaumont, M., Jackson, M.A., Steves, C.J., Spector, T.D., and Bell, J.T. (2018). Heritable components of the human fecal microbiome are associated with visceral fat. Gut Microb. *9*, 61–67. https://doi.org/10.1080/19490976.2017.1356556.

48. Beaumont, M., Goodrich, J.K., Jackson, M.A., Yet, I., Davenport, E.R., Vieira-Silva, S., Debelius, J., Pallister, T., Mangino, M., Raes, J., et al. (2016). Heritable components of the human fecal microbiome are associated with visceral fat. Genome Biol. *17*, 189. https://doi.org/10.1186/s13059-016-1052-7.

49. Benítez-Páez, A., Gómez Del Pugar, E.M., López-Almela, I., Moya-Pérez, Á., Codoñer-Franch, P., and Sanz, Y. (2020). Depletion of blautia species in the microbiota of obese children relates to intestinal inflammation and metabolic phenotype worsening. mSystems *5*, 00857-19. https://doi.org/10.1128/mSystems.00857-19.

50. Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J.M., Kennedy, S., et al. (2013). Richness of human gut microbiome correlates with metabolic markers. Nature *500*, 541–546. https://doi.org/10.1038/nature12506.

51. Porras, D., Nistal, E., Martínez-Flórez, S., González-Gallego, J., García-Mediavilla, M.V., and Sánchez-Campos, S. (2018). Intestinal microbiota modulation in obesity-related non-alcoholic fatty liver disease. Front. Physiol. *9*, 1813. https://doi.org/10.3389/fphys.2018.01813.

52. Castaner, O., Goday, A., Park, Y.M., Lee, S.H., Magkos, F., Shiow, S.A.T.E., and Schröder, H. (2018). The gut microbiome profile in obesity: a systematic review. Int. J. Endocrinol. *2018*, 4095789. https://doi.org/10.1155/2018/4095789.

53. Thingholm, L.B., Rühlemann, M.C., Koch, M., Fuqua, B., Laucke, G., Boehm, R., Bang, C., Franzosa, E.A., Hübenthal, M., Rahnavard, A., et al. (2019). Obese individuals with and without type 2 diabetes show different gut microbial functional capacity and composition. Cell Host Microbe *26*, 252–264.e10. https://doi.org/10.1016/j.chom.2019.07.004.

54. Rautio, M., Eerola, E., Väisänen-Tunkelrott, M.L., Molitoris, D., Lawson, P., Collins, M.D., and Jousimies-Somer, H. (2003). Reclassification of bacteroides putredinis (weinberg et al., 1937) in a new genus alistipes gen. Nov., as alistipes putredinis comb. Nov., and description of alistipes finegoldii sp. Nov., from human sources. Syst. Appl. Microbiol. *26*, 182–188. https://doi.org/10.1078/072320203322346029.

55. Cheng, C.W., Lin, H.S., Ye, J.J., Yang, C.C., Chiang, P.C., Wu, T.S., and Lee, M.H. (2009). Clinical significance of and outcomes for bacteroides fragilis bacteremia. J. Microbiol. Immunol. Infect. *42*, 243–250.

56. Parker, B.J., Wearsch, P.A., Veloo, A.C.M., and Rodriguez-Palacios, A. (2020). The genus alistipes: gut bacteria with emerging implications to inflammation, cancer, and mental health. Front. Immunol. *11*, 906. https://doi.org/10.3389/fimmu.2020.00906.

57. Wexler, H.M. (2007). Bacteroides: the good, the bad, and the nitty-gritty. Clin. Microbiol. Rev. *20*, 593–621. https://doi.org/10.1128/CMR.00008-07.

58. Brahe, L.K., Le Chatelier, E., Prifti, E., Pons, N., Kennedy, S., Hansen, T., Pedersen, O., Astrup, A., Ehrlich, S.D., and Larsen, L.H. (2015). Specific gut microbiota features and metabolic markers in postmenopausal women with obesity. Nutr. Diabetes *5*, e159. https://doi.org/10.1038/nutd.2015.9.

59. Lim, M.Y., You, H.J., Yoon, H.S., Kwon, B., Lee, J.Y., Lee, S., Song, Y.M., Lee, K., Sung, J., and Ko, G. (2017). The effect of heritability and host genetics on the gut microbiota and metabolic syndrome. Gut *66*, 1031–1038. https://doi.org/10.1136/gutjnl-2015-311326.

60. Bian, X., Wu, W., Yang, L., Lv, L., Wang, Q., Li, Y., Ye, J., Fang, D., Wu, J., Jiang, X., et al. (2019). Administration of akkermansia muciniphila ameliorates dextran sulfate sodium-induced ulcerative colitis in mice. Front. Microbiol. *10*, 2259. https://doi.org/10.3389/fmicb.2019.02259.

61. Hiippala, K., Barreto, G., Burrello, C., Diaz-Basabe, A., Suutarinen, M., Kainulainen, V., Bowers, J.R., Lemmer, D., Engelthaler, D.M., Eklund, K.K., et al. (2020). Novel odoribacter splanchnicus strain and its outer membrane vesicles exert immunoregulatory effects in vitro. Front. Microbiol. *11*, 575455. https://doi.org/10.3389/fmicb.2020.575455.

62. Bodogai, M., O'Connell, J., Kim, K., Kim, Y., Moritoh, K., Chen, C., Gusev, F., Vaughan, K., Shulzhenko, N., Mattison, J.A., et al. (2018). Commensal bacteria contribute to insulin resistance in aging by activating innate b1a cells. Sci. Transl. Med. *10*, eaat4271. https://doi.org/10.1126/scitranslmed.aat4271.

63. Frankel, A.E., Deshmukh, S., Reddy, A., Lightcap, J., Hayes, M., McClellan, S., Singh, S., Rabideau, B., Glover, T.G., Roberts, B., et al. (2019). Cancer immune checkpoint inhibitor therapy and the gut microbiota. Integr. Cancer Ther. *18*. 1534735419846379. https://doi.org/10.1177/1534735419846379.

64. Bowyer, R.C.E., Jackson, M.A., Le Roy, C.I., Ni Lochlainn, M., Spector, T.D., Dowd, J.B., and Steves, C.J. (2019). Socioeconomic status and the gut microbiome: a twinsuk cohort study. Microorganisms *7*, 17. https://doi.org/10.3390/microorganisms7010017.

65. Melgar-Lesmes, P., Garcia-Polite, F., Del-Rey-Puech, P., Rosas, E., Dreyfuss, J.L., Montell, E., Vergés, J., Edelman, E.R., and Balcells, M. (2016). Treatment with chondroitin sulfate to modulate inflammation and atherogenesis in obesity. Atherosclerosis *245*, 82–87. https://doi.org/10.1016/j.atherosclerosis.2015.12.016.

66. Visekruna, A., and Luu, M. (2021). The role of short-chain fatty acids and bile acids in intestinal and liver function, inflammation, and carcinogenesis. Front. Cell Dev. Biol. *9*, 703218. https://doi.org/10.3389/fcell.2021.703218.

67. Zeng, H., Umar, S., Rust, B., Lazarova, D., and Bordonaro, M. (2019). Secondary bile acids and short chain fatty acids in the colon: a focus on colonic microbiome, cell proliferation, inflammation, and cancer. Int. J. Mol. Sci. *20*, 1214. https://doi.org/10.3390/ijms20051214.

68. Cai, S.Y., Ouyang, X., Chen, Y., Soroka, C.J., Wang, J., Mennone, A., Wang, Y., Mehal, W.Z., Jain, D., and Boyer, J.L. (2017). Bile acids initiate cholestatic liver injury by triggering a hepatocyte-specific inflammatory response. JCI Insight *2*, e90780. https://doi.org/10.1172/jci.insight.90780.

69. Ocvirk, S., and O'Keefe, S.J. (2017). Influence of bile acids on colorectal cancer risk: potential mechanisms mediated by diet - gut microbiota interactions. Curr. Nutr. Rep. *6*, 315–322. https://doi.org/10.1007/s13668-017-0219-5.

70. Zatterale, F., Longo, M., Naderi, J., Raciti, G.A., Desiderio, A., Miele, C., and Beguinot, F. (2019). Chronic adipose tissue inflammation linking obesity to insulin resistance and type 2 diabetes. Front. Physiol. *10*, 1607. https://doi.org/10.3389/fphys.2019.01607.

71. Duan, Y., Zeng, L., Zheng, C., Song, B., Li, F., Kong, X., and Xu, K. (2018). Inflammatory links between high fat diets and diseases. Front. Immunol. *9*, 2649. https://doi.org/10.3389/fimmu.2018.02649.

72. Yang, J., Li, Y., Wen, Z., Liu, W., Meng, L., and Huang, H. (2021). Oscillospira - a candidate for the next-generation probiotics. Gut Microb. *13*, 1987783. https://doi.org/10.1080/19490976.2021.1987783.

73. Schneeberger, M., Everard, A., Gómez-Valadés, A.G., Matamoros, S., Ramírez, S., Delzenne, N.M., Gomis, R., Claret, M., and Cani, P.D. (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. https://doi.org/10.1038/srep16643.

74. Rodríguez-Carrio, J., Salazar, N., Margolles, A., González, S., Gueimonde, M., de Los Reyes-Gavilán, C.G., and Suárez, A. (2017). Free fatty acids profiles are related to gut microbiota signatures and short-chain fatty

acids. Front. Immunol. *8*, 823. https://doi.org/10.3389/fimmu.2017.00823.

75. Garner, M.R., Flint, J.F., and Russell, J.B. (2002). Allisonella histaminiformans gen. Nov., sp nov. A novel bacterium that produces histamine, utilizes histicline as its sole energy source, and could play a role in bovine and equine laminitis. Syst. Appl. Microbiol. *25*, 498–506. https://doi.org/10.1078/07232020260517625.

76. Jiang, P., Wu, S., Luo, Q., Zhao, X.M., and Chen, W.H. (2021). Metagenomic analysis of common intestinal diseases reveals relationships among microbial signatures and powers multidisease diagnostic models. mSystems *6*, e00112-21. https://doi.org/10.1128/mSystems.00112-21.

77. Wirbel, J., Pyl, P.T., Kartal, E., Zych, K., Kashani, A., Milanese, A., Fleck, J.S., Voigt, A.Y., Palleja, A., Ponnudurai, R., et al. (2019). Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. Nat. Med. *25*, 679–689. https://doi.org/10.1038/s41591-019-0406-6.

78. Pigeyre, M., Yazdi, F.T., Kaur, Y., and Meyre, D. (2016). Recent progress in genetics, epigenetics and metagenomics unveils the pathophysiology of human obesity. Clin. Sci. *130*, 943–986. https://doi.org/10.1042/CS20160136.

79. Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. *37*, 852–857. https://doi.org/10.1038/s41587-019-0209-9.

80. Douglas, G.M., Maffei, V.J., Zaneveld, J.R., Yurgel, S.N., Brown, J.R., Taylor, C.M., Huttenhower, C., and Langille, M.G.I. (2020). PICRUSt2 for prediction of metagenome functions. Nat. Biotechnol. *38*, 685–688. https://doi.org/10.1038/s41587-020-0548-6.

81. Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., and Holmes, S.P. (2016). Dada2: High-resolution sample inference from illumina amplicon data. Nat Methods *13*, 581–583. https://doi.org/10.1038/nmeth.3869.

82. Dixon, P. (2003). Vegan, a package of r functions for community ecology. J. Veg. Sci. *14*, 927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x.

83. Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., and Tiedje, J.M. (2014). Ribosomal database project: data and tools for high throughput rrna analysis. Nucleic Acids Res. *42*, D633–D642. https://doi.org/10.1093/nar/gkt1244.

84. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

85. Beghini, F., McIver, L.J., Blanco-Miguez, A., Dubois, L., Asnicar, F., Maharjan, S., Mailyan, A., Manghi, P., Scholz, M., Thomas, A.M., et al. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. Elife *10*, e65088. https://doi.org/10.7554/eLife.65088.

86. Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C.A., Holland, T.A., Keseler, I.M., Kothari, A., Kubo, A., et al. (2014). The MetaCyc database of metabolic pathways and enzymes and the biocyc collection of pathway/genome databases. Nucleic Acids Res. *42*, D459–D471. https://doi.org/10.1093/nar/gkt1103.

87. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849. https://doi.org/10.1093/bioinformatics/btw313.

88. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. J. Mach. Learn. Res. *12*, 2825–2830.

89. Chen, T., and Guestrin, C. (2016). XGBoost: a scalable tree boosting system. In 22nd SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

90. Sugiyama, M., Suzuki, T., Nakajima, S., Kashima, H., von Bünau, P., and Kawanabe, M. (2008). Direct importance estimation for covariate shift adaptation. Ann. Inst. Stat. Math. *60*, 699–746. https://doi.org/10.1007/s10463-008-0197-x.

91. Yu, L., Zhilu, X., Yun Kit, Y., Hein, M.T., Wenli, H., Wei, J., Francis, K.L.C., and Siew, C.N. (2022). Meta-analysis of Gut Microbiome Studies Identifies Consistent Obesity-specific Microbial Signatures and Shared Metabolic Pathways (Open Science Framework). https://doi.org/10.17605/OSF.IO/5TNBU.

92. Moher, D., Liberati, A., Tetzlaff, J., and Altman, D.G.; PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ *339*, b2535. https://doi.org/10.1136/bmj.b2535.

93. Purnell, J.Q. (2000). Definitions, Classification, and Epidemiology of Obesity (Endotext).

94. World Health Organization; Regional Office for the Western, Pacific (2000). The Asia-Pacific Perspective: Redefining Obesity and its Treatment (Health Communications Australia).

95. Vanderwall, C., Eickhoff, J., Randall Clark, R., and Carrel, A.L. (2018). BMI z-score in obese children is a poor predictor of adiposity changes over time. BMC Pediatr. *18*, 187. https://doi.org/10.1186/s12887-018-1160-5.

96. Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen, E., Knight, R., Huttley, G.A., and Gregory Caporaso, J. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome *6*, 90. https://doi.org/10.1186/s40168-018-0470-z.

97. Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.H., Whitman, W.B., Euzéby, J., Amann, R., and Rosselló-Móra, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16s rrna gene sequences. Nat. Rev. Microbiol. *12*, 635–645. https://doi.org/10.1038/nrmicro3330.

98. Martín-Fernández, J.A., Hron, K., Templ, M., Filzmoser, P., and Palarea-Albaladejo, J. (2015). Bayesian-multiplicative treatment of count zeros in compositional data sets. Stat. Model. Int. J. *15*, 134–158. https://doi.org/10.1177/1471082x14535524.

99. Hinton, G.E. (1989). Connectionist learning procedures. Artif. Intell. *40*, 185–234. https://doi.org/10.1016/0004-3702(89)90049-0.

100. Breiman, L. (2001). Random forests. Mach. Learn. *45*, 5–32. https://doi.org/10.1023/A:1010933404324.

101. Huang, G.B. (2003). Learning capability and storage capacity of two-hidden-layer feedforward networks. IEEE Trans. Neural Netw. *14*, 274–281. https://doi.org/10.1109/TNN.2003.809401.

102. Shibata, K., and Yusuke, I. (2009). Effect of number of hidden neurons on learning in large-scale layered neural networks. ICCAS-SICE *2009*, 5008–5013.

103. Trenn, S. (2008). Multilayer perceptrons: approximation order and necessary number of hidden units. IEEE Trans. Neural Netw. *19*, 836–844. https://doi.org/10.1109/TNN.2007.912306.

104. Xu, S., and Chen, L. (2008). A novel approach for determining the optimal number of hidden layer neurons for FNN's and its application in data mining. ICITA *2008*, 23–26.

105. Shimodaira, H. (2000). Improving predictive inference under covariate shift by weighting the log-likelihood function. J. Stat. Plan. Inference *90*, 227–244. https://doi.org/10.1016/S0378-3758(00)00115-4.

106. Pereira, J.P.B., Stroes, E.S.G., Zwinderman, A.H., and Levin, E. (2022). Covered information disentanglement: model transparency via unbiased permutation importance. In Proceedings of the AAAI Conference on Artificial Intelligence, *36*, pp. 7984–7992.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Ahmad et al.[23]	SRA	PRJNA554535
Chavez et al.[24]	SRA	PRJNA417691
Barengolts et al.[25]	SRA	PRJNA389481
Gao et al.[26]	SRA	PRJNA486071
Olsson et al.[27]	SRA	PRJEB33908
Lippert et al.[28]	SRA	PRJNA339677
Ross et al.[29]	SRA	SRP053023
Zupancic et al.[30]	dbGap	phs000258.v2.p1
Escobar et al.[31]	SRA	ERP003466
Wu et al.[32]	SRA	SRX089367
Zeevi et al.[33]	SRA	PRJEB11532
Goodrich et al.[34]	SRA	PRJEB6705
Baxter et al.[35]	SRA	SRP062005
HKGutMicMap[36]	SRA	PRJNA557323/PRJNA815750
He et al.[37]	ENA	PRJEB18535
HMP et al.[38]	dbGap	phs000228.v4.p1
AGP et al.[39]	SRA	PRJEB11419
Stanislawski et al.[40]	SRA	ERP112799
Software and algorithms		
QIIME 2 2020.11	Bolyen et al.[79]	docs.qiime2.org/2020.11/
PICRUSt V2.4.1	Douglas et al.[80]	github.com/picrust/picrust2
DADA2 2020.11.1	Callahan et al.[81]	docs.qiime2.org/2020.11/plugins/available/dada2/?highlight=dada2
Vegan V2.6-2	Dixon et al.[82]	github.com/vegandevs/vegan
RDP 11.5	Cole et al.[83]	sourceforge.net/projects/rdp-classifier/
Trimmomatic 0.39	Bolger et al.[84]	www.usadellab.org/cms/?page=trimmomatic
KneadData 0.10.0	Github	github.com/biobakery/kneaddata
MetaPhlAn 3.0.10	Francesco et al.[85]	github.com/biobakery/MetaPhlAn
MetaCyc V26.1	Caspi et al.[86]	metacyc.org/
ComplexHeatmap V2.13.1	Gu et al.[87]	github.com/jokergoo/ComplexHeatmap
sklearn 1.0.2	Pedregosa et al.[88]	scikit-learn.org/stable/
xgboost 0.90	Chen et al.[89]	github.com/dmlc/xgboost
KLIEP function	Sugiyama et al.[90]	github.com/srome/pykliep
R 3.6.3	CRAN	r-project.org
Python 3.7.0	python.org	www.python.org/downloads/release/python-370/
Other		
Resource website for this paper	This paper	github.com/MetagenomicsYL/meta_obesity

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by Siew Chien Ng (siewchienng@cuhk.edu.hk).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The 16S rRNA sequencing data and metadata of HKGutMicMap are available under the accession number of PRJNA815750 in the SRA database and are publicly available as of the date of publication. Other studys' sequencing data and metadata can be downloaded from ENA, NCBI SRA or NCBI dbGap database. The detailed accession numbers of each dataset are listed in Table S1.
- The key source codes that supported these results and findings can be found on GitHub (https://github.com/MetagenomicsYL/meta_obesity) and is publicly available as of the date of publication.
- Any additional information required to re-analyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Literature search

This meta-analysis was conducted according to a preregistered protocol[91] and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline.[92] We reviewed literature from Medline and Ovid databases for research papers published between January 2010 and January 2022. Keywords, including "obese", "obesity", "overweight", "body mass index", "BMI", "microbiota", "microbiome", "fecal", "fecal", "gut", "intestinal", were used to search for eligible articles. The search strategies used to identify potential studies are described in detail in Figure S1. Studies were included if they (1) used fecal samples for 16S sequencing; (2) had a cross-sectional design consisting of obese individuals and healthy lean controls (BMI larger than 25 or 30 was considered as obese cases in these studies, as the definition of obesity may vary in Western and Asian populations), and (3) had publicly available sequence and subject metadata. Studies were excluded if they (1) assessed specific microbiota of patients with intervention by supplementation or therapy, (2) did not include healthy lean controls, (3) were clinical trials for probiotics or other diet-related treatments, (4) were comprised of participants diagnosed with inflammatory bowel disease (IBD), *Clostridioides difficile* infection (CDI), or colorectal cancer (CRC).

For data requiring a research plan for access, we submitted research proposals and allowed 30 days for approval, after which we excluded the dataset if no approvals were given. The literature search was conducted independently by four authors and any disagreements during the inclusion/exclusion process were resolved through discussion among the four researchers. Both adults and children over the age of three were included in the study population in accordance with the inclusion criteria of Stanislawski's meta-analysis.[22] Research papers describing gut microbiota surveys in general populations with accompanying BMI information were included as the previous meta-analysis published by Sze et al. also covered such data.[21] In microbiome population surveys with BMI information, we chose individuals with BMI >30 kg/m^2 as obese and those with BMI <25 kg/m^2 as lean in Western populations to maximize differences in gut microbiota composition.[93] For Asian populations, we considered BMI >25 kg/m^2 as obese and BMI <23 kg/m^2 as healthy lean controls, consistent with the World Health Organization definition of obesity in Asia.[94] For children populations <18 years old, we followed obesity criteria defined by the US Centers for Disease Control and Prevention, in which a child with BMI *Z*score >1.64 is considered obese and BMI *Z*score <1.04 considered lean.[95]

Bacterial community diversity analysis

Raw gut microbiota 16S rRNA gene sequencing datasets were downloaded from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) or European Nucleotide Archive (ENA) and independently processed using QIIME 2[79] to obtain the taxonomy annotation. Low-quality bases were trimmed and sequences were denoised and chimera filtered using DADA2.[81] The taxonomy of each ASV was assigned using the q2 feature classifier[96] trained with the Naive Bayes algorithm on the RDP 11.5 database (https://sourceforge.net/projects/rdp-classifier/).[83] Low frequency ASVs detected in fewer than three samples or with a total count of less than 10 sequences were removed. Counts belonging to ASVs with identical taxonomic annotations were summed in the final feature table. Samples with a total feature count of less than 1000 sequences were excluded from each study. For functional profiling, we used Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2)[80] to

predict functional pathways using feature tables derived from QIIME 2. Functional pathways were inferred for each ASV based on the MetaCyc database (https://metacyc.org/).[86] To summarize functional information at broader levels, we categorized functional pathways into Superclass levels according to the MetaCyc database.

To investigate differences in microbial diversity and F/B ratio in obese and healthy controls, alpha diversity and beta diversity were calculated for all studies. Feature counts of individual samples were subsampled according to the lowest sample depth in each study. Commonly used alpha diversity metrics including Shannon index, Evenness, Chao1 index and Observed OTUs (OBS) were computed using diversity, estimateR and specnumber function from the vegan package, respectively.[82] The statistical significance of each metric in each study was identified using Welch's t-test. For beta diversity, Bray-Curtis dissimilarities were computed for all studies and visualized with Principal Coordinate Analysis (PCoA) using the vegan package.[82] ANOSIM tests with 999 permutations were applied to assess statistical differences in microbial community composition between groups. The F/B ratio was calculated by aggregating the feature table based on the phylum level taxonomy and dividing the abundance of Firmicutes by Bacteroidetes in each sample. To identify differences in alpha diversity and F/B ratios in obese vs. healthy individuals, a $\log_2$ transformation was applied and the $\log_2$ fold change between obese and healthy groups was computed.

There were many sources of variation such as DNA extraction method, sequencing platform, geographical regions, host genetic backgrounds, age, gender, etc, that might contribute to the high variance in the microbiome composition. To take account of these variations from different studies, we applied the random effects model, also known as Generalized Linear Mixed Models (GLMM), to identify signatures for obesity with the study as a random effect. The race-specific, age-specific, and gender-specific variation in the gut microbiome was also taken into consideration in this study using the subgroup analysis. To estimate the effect of obesity on alpha diversity using combined studies, we applied random effects models to calculate the combined $\log_2$ fold change of each diversity metric and F/B ratio between obese and healthy groups. The study heterogeneity was identified using the Higgins & Thompson' I^2 statistics and the publication bias was evaluated by the Funnel plot with Egger's regression test. As there were several studies not reporting the antibiotics usage of participants, the sensitivity analysis was conducted and applicable studies were excluded to eliminate the potential impact of antibiotics on the combined alpha diversity and F/B ratio. In addition, we also evaluated correlations between BMI and alpha diversity using random effects models for datasets with BMI information. To identify relationships between BMI and diversity across different ethnicities, age groups and sex, subjects were separated into different subgroups and correlation between diversity and BMI was evaluated.

QUANTIFICATION AND STATISTICAL ANALYSIS

Differential abundance analysis of taxa and functional modules

As short length 16S rRNA gene sequencing data are more likely to be reliable at the genus level,[97] all ASVs were summarized at the genus level according to inferred taxonomy. Genera and functional modules detected in more than 2/3 studies were used for differential abundance analysis. The $\log_2$ fold changes of genera between obese and healthy groups were calculated and correlations between these taxa and BMI were evaluated in the combined studies. To generate log-ratio relative abundances in the microbial composition data, we first transformed data by replacing zero values using Bayesian-Multiplicative replacement of count zeros as log-ratio transformations require data with positive numbers.[98] Random effects models and Stratified Wilcoxon tests were applied to estimate statistical difference of genera and functional modules. p-values derived from Stratified Wilcoxon tests were adjusted for multiple testing using the Benjamini-Hochberg method to estimate false discovery rate (FDR). For subgroup analysis, we stratified the population according to their ethnic group, age group or sex and assessed relationships between taxa abundance and BMI. The mean $\log_{10}$ abundance of functional modules with $p < 0.05$ in the random effects model and FDR < 0.01 in the Stratified Wilcoxon tests was calculated in the obese and healthy groups. In addition, MetaCyc Superclass level functional modules were compared between obese and healthy individuals using Wilcoxon rank sum tests. Correlations between genera and functional pathways were evaluated and visualized as heatmaps using the ComplexHeatmap package[87] in R.

Species identification using shotgun metagenome data

During the literature search process, we found that there was a limited number of obesity microbiome association studies providing publicly available shotgun metagenomic data. To identify the species-level

markers, we chose to extend our meta-analysis result in the included studies which contained both shotgun metagenomic and 16S sequence data, rather than synthesizing additional shotgun metagenomic datasets. One of the included studies from a Hong Kong gut microbiota population survey had corresponding shotgun metagenomic sequence data. We therefore downloaded the raw metagenomes to infer species-level community compositions, as opposed to genera for 16S data. Trimmomatic and KneadData were used for trimming paired-end sequencing data with low quality scores and removing human contamination.[84] Microbial community composition was inferred from quality filtered sequence data using Metagenomic Phylogenetic Analysis (MetaPhlAn3).[85] Candidate species were selected for further analyses based on the list of implicated genera identified from 16S data. Species that significantly differed between groups and correlated with BMI were identified as species-level candidates of obesity.

Machine learning models for obesity classification

To examine whether the obese-related microbiome could reliably discriminate the obesity phenotype, we used machine learning models including XGBoost,[89] neural network,[99] and random forest,[100] to build classification models to identify obese individuals and healthy controls. The model with the best classification ability for obesity was applied for further analysis. All microbial genera were used to build machine learning model. We applied three strategies for model training: (1) models were trained within each study with 75% of the data as training set and the remaining 25% as validation, (2) 10-fold cross-validation using combined data, (3) leave-one-study-out validation. Model performance was evaluated using AUC and the permutation test for each AUC was reported in the heatmap. The dependent variables were permuted 100 times in the permutation test for each AUC.

We identified the optimal parameter sets for each model using the grid search approach. For hyperparameter tuning in the XGBoost model, we randomly tried several parameters that control the model complexity and parameters that add randomness to the model. We found that n_estimators, max_depth, and learning_rate had a larger impact on the model performance than other parameters. Thus, the grid search was performed to identify the best combination of n_estimators, max_depth and learning_rate for the XGBoost model. The values we tested for the n_estimators were: {100, 200, 300, 400, 500, 600, 700, 800, 900, 1000}, for the max_depth were: {3, 4, 5, 6, 7, 8, 9, 10}, and for learning_rate were {0.05, 0.1, 0.25, 0.5, 1.0}. The best combination of parameters {n_estimators = 1000, max_depth = 4, learning_rate = 1} was applied to build the XGBoost model using populations from 18 cohorts for 10-fold cross-validation.

To optimize the parameter in the random forest model, we found the parameters, max_depth, n_estimators and max_features, had a large influence on the model performance. The values we tested for the n_estimators were: {100, 200, 300, 400, 500, 600, 700, 800, 900, 1000}, for the max_depth were: {3, 4, 5, 6, 7, 8, 9, 10}, and for max_features were {sqrt, log2, None}. We used grid search to test different combinations of these parameters and applied {n_estimators =100, max_depth = 10, max_features = None} to the random forest model training for 10-fold cross-validation.

To identify the optimal parameter sets for the neural network, the grid search was applied for parameter optimization. The values for grid search were hidden_layer_sizes: {(258, 86), (126), (94), (18)}, activation: {tanh, relu}, solver: {sgd, adam}, alpha: {0.0001, 0.001, 0.01}, learning_rate: {constant, adaptive}. The hidden_layer_sizes were selected according to several formulas for the neural networks.[101–104] The parameter {hidden_layer_sizes = (258, 86), activation = relu, solver = adam, alpha = 0.0001, learning_rate = constant} was applied to the neural network model training for 10-fold cross-validation. For leave-one-study-out validation and models trained within each study, each model was tuned parameters by the grid search using the same range of hyperparameters mentioned above.

In this meta-analysis, we combined study cohorts from various ethnic backgrounds and different geographic regions. High variations of microbiome structure were observed from study cohorts. A big challenge for the machine learning classification is that the training data distribution differs from the test one, called covariate shift. The impact of covariate shift can be reduced by weighting the log likelihood terms according to the importance, which is the ratio of testing and training samples density.[90,105] We used the Kullback-Leibler Importance Estimation Procedure (KLIEP),[85] a commonly used covariate shift adaption method, to address the covariate shift issue in the dataset of combined cohorts. In a more computationally efficient manner, the KLIEP is capable of calculating the importance directly without the need for density

estimation, and the importance can be used as sample weights for the training set when the machine learning model is being trained. The identification of permutation importance of features is a popular way to explain model predictions.[106] The permutation feature importance of each microbial feature was calculated according to the decrease in the AUC of the machine learning model when the microbial feature is randomly shuffled.[100] The key source codes that supported these results and findings can be found at GitHub (https://github.com/MetagenomicsYL/meta_obesity).