THE ROLE OF GUT MICROBIOME AND HOST METABOLOME IN DIET AND CARDIOVASCULAR RISK FACTORS

Yiqing Wang

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Nutrition in the Gillings School of Global Public Health.

Chapel Hill
2020

Approved by:

Penny Gorden-Larsen

Annie Green Howard

Christy Avery

Katie Meyer

Susan Sumner

ABSTRACT

Yiqing Wang: The Role of Gut Microbiome and Host Metabolome in Diet and Cardiovascular Risk Factors
(Under the direction of Penny Gordon-Larsen)

High blood pressure (BP) and high adiposity are leading risk factors for cardiovascular morbidity and mortality. Excessive sodium, low potassium, and low fiber intake are among the main contributors to elevated BP and adiposity. We hypothesize that the gut microbiota and host metabolites are influenced by diet and have effects on cardiovascular health. Numerous animal model studies have suggested that dietary intake impacts host BP and adiposity particularly through microbiota-mediated short-chain fatty acids (SCFAs). Yet, there is limited empirical evidence in humans to inform current understanding.

To address this research gap, we used a cross-sectional adult cohort from the population-based Chinese Health and Nutrition Survey (CHNS) with rich gut microbiota (16S rRNA) and plasma metabolomics data, high-quality diet data from three-consecutive 24-h validated recalls and household food inventories, and health data from physical examinations. We first examined the association between sodium and potassium with gut microbiota and plasma metabolites. We then investigated the associations between gut microbiota and plasma metabolites with systolic (SBP) and diastolic BP (DBP). Last, we assessed the associations between plasma SCFAs and two adiposity measures, body mass index (BMI) and waist-to-height ratio (WHtR).

After accounting for geographic variation in microbiota, we found that that sodium and potassium consumption was associated with microbiota and metabolites, including Staphylococcus, microbiota-derived phenolics and SCFAs previously linked to inflammation, hypertension and adiposity. While we did not observe statistical evidence for an association between SCFAs and BP, we found that

sphingomyelins, acyl-carnitines, and a lipids pattern characterized by long-chain fatty acids were positively associated with BP. We also found positive associations between SCFAs with BMI and WHtR.

Our results suggest that gut microbiota and related metabolites may play an important role in the dietary etiology of cardiovascular disease (CVD). As such, our findings provide insights into potential dietary interventions targeting microbiota or metabolites for disease prevention and treatment. Future longitudinal and randomized-controlled studies are needed to determine the causal relationships between diet, gut microbiome, host metabolome, and CVD risk factors.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor and mentor, Dr. Penny Gordon-Larsen, for providing me with resources I needed and supporting me with patience and reassurance over the last four years. Penny could always guide me through the thought process when I seem to have exhausted all ideas and give thoughtful feedbacks. Her trust in my abilities has helped me transition from a shy student to a confident researcher who is not afraid of speaking up her mind and presenting her work. Without her guidance and encouragement, I would not have the confidence to pursue a career in scientific research. Lastly but not least, I would like to thank Penny for giving me freedom to learn and try different analytical skills, considering my career and personal development, introducing me to her colleagues and friends, and helping me secure a great postdoctoral position.

I am also grateful for my excellent PhD committee, including Annie Green Howard, who has helped me get familar with our study cohort, supervised my statistical analyses, and met with me on a weekly basis from the beginning of my study; Christy Avery, who led me into the world of cardiovascular and genetic epidemiology and always reminded me not to forget to think critically; Katie Meyer, who provided careful thought about my research plan and helpful advice for diet and microbiome data analyses; and Susan Sumner, who never hesitated to offer support and resources to me to learn about metabolomics pipeline and analyses.

Anthony Fodor and his postdoc, Shan Sun, of the Department of Bioinformatics at University of North Carolina at Charlotte, though not members of my thesis committee, provided indispensable guidance and valuable advice for my dissertation work, particular for the gut microbiome analyses. His former student, Matthew Tsilimigras, now a postdoc of Penny's group, also provided tremendous technical and thoughtful advice to this work. I thank them for their patience and all of their help. I also

want to thank Misa Graff from the Department of Epidemiology, who generously shared her programs for making polygenetic risk scores and spent countless hours help me trouble-shooting the code, although this work did not make it to the final dissertation.

This work was made possible by the Carolina Population Center and the China Health and Nutrition Surevey (CHNS) group, especially Shufa Du, who oversighted the data collection, and Hsiao-Chuan Tein and Guifeng Jin, who managed the datasets used in this dissertation. I am grateful to the Department of Nutrition faculty, especially Linda Adair and Barry Popkin, who provided immense support for my early work and professional mentorship, as well as the lead academic coordinator, Jonathan Earnest, for his kindness considerations during the hardtime of quarantine. I am also thankful for my peers for their support and kind friendship, including Melissa Jensen, Ruixue Hou, Allison Lacko, Laetitia Meyrueix, Natalia Rebolledo, Xinruo Zhang, Tania Aburto, Charlotte Lane, Mike Essman, Luis Maldonado, William Green, and Khristopher Nicholas.

I am grateful for the unconditional love and support of my family, especially my mom who traveled from China to North Carolina to visit me, and my husband, Boyu Zhang, who is always able to be calm me down during my moments of self-doubt and anxiety. Withour their support, nothing I accomplished so far would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ARIC	Atherosclerosis Risk in Communties
BMI	Body mass index
BP	Blood pressure
CARDIA	Coronary Artery Risk Development in Young Adults
CHNS	China Health and Nutrition Survey
CI	Confidence interval
CNTCS	China Nutritional Transition Cohort Study
CVD	Cardiovascular diseases
DASH	Dietary Approaches to Stop Hypertension
DBP	Diastolic blood pressure
DMTPA	2,3-dihydroxy-5-methylthio-4-pentenoate
eGFR	Estimated glomerular filtration rate
FCT	Food composition table
FDR	False discovery rate
GALFs	Glycerrhetinic acid like factors
dbRDA	Distance-based redundancy analysis
GPCRs	G protein-coupled receptors
MDS	Multiple dimension scaling
Na/K ratio	Sodium to potassium ratio
NSAIDs	Non-steroid anti-inflammatory drugs
OTUs	Operational Taxonomic Units
PC	Phosphatidylcholine
PCA	Principle component analysis
PCoA	Principle coordinate analysis

PERMANOVA	Permutational multivariate analysis of variance
PI	Phosphatidylinositol
PPIs	Proton pump inhibitors
RMSE	Root mean square error
SBP	Systolic blood pressure
SCFA	Short chain fatty acids
SD	Standard deviation
T_H17	T helper 17
UK	United Kindom
US	United States
WHO	World Health Organization
WHtR	Waist-to-height ratio
5x2cv	5 iterations of 2-folds cross-validation

CHAPTER 1. INTRODUCTION

Background

The prevalence of the two main cardiovascular disease risk factors, elevated adiposity and blood pressure, have increased dramatically around the world over the past decades. In particular, China has the highest absolute burden of elevated blood pressure and has even greater increase in abdominal adiposity relative to the overall body mass, as measured by waist circumference and body mass index (BMI), respectively. Given that diet affects the intestinal bacteria and the metabolites of which subsequently impact the host physiology, it is hypothesized that the gut microbiota may mediate or modify the diet-health relationship. For example, dietary carbohydrate feeds the gut microbiota that produce short-chain fatty acids (SCFAs), which is absorbed into the portal system to affect blood pressure and fat oxidation. Although numerous experimental studies using animal models have provided the biological basis for these complex relationships, there is a lack of empirical evidence in humans that could potentially apply to the general population. In addition, some of the observational findings are inconsistent, owing to different populations, small sample sizes, small variations in diet and/or health outcomes, and insufficient adjustment of potential confounding.

In addressing these research gaps, we capitalized on the population-based Chinese Health and Nutrition Survey (CHNS). The CHNS is a population-based prospective cohort across 12 provinces and three megacities in China that includes rich gut microbiota 16S rRNA data from 3,408 adults aged 18-80 years in 2015, with plasma metabolomics (n=400) and gut whole metagenome data (n=214) collected from sub-samples. Additionally, high-quality diet data from three-consecutive validated individual 24-h recalls and household food inventories, health data from physical examinations, as well as sociodemographic data were collected by trained examiners. The CHNS participants were characterized

by diverse urban and rural diet, habitual consumption of excessive sodium and deficient potassium, and low treatment rate for hypertension, making them an ideal population examine the associations between key dietary and health risk factors for cardiovascular diseases like sodium and blood pressure.

Unsing this unique cohort, we aimed to investigate (1) the associations between dietary sodium and potassium consumption with gut microbiota and plasma metabolites; (2) the associations between gut microbiota and plasma metabolites with systolic and diastolic blood pressure (SBP and DBP); (3) the association between plasma SCFAs and two adiposity measure, BMI and waist-to-height ratio (WHtR), and whether the gut microbiota of subjects with higher BMI and/or WHtR had higher capacity to harvest energy through carbohydrate fermentation, which could be partly reflected by the levels of circulating SCFAs.

Specific aims

The primary objective of the current research was to improve our understanding of the underlying mechanisms of gut microbial and related metabolic pathways in the relationship between diet and cardiovascular risk factors, particularly in regard to key dietary risk factors, SCFAs, blood pressure, and adiposity measures. We sought to achieve this objective with the following specific aims:

Aim 1. Examine the association between sodium and potassium consumption with gut microbiota (16S rRNA) and plasma metabolites.

a. Use multivariable-adjusted linear regression to analyze microbiota within-person diversity (α-diversity) and specific taxa, and distance-based redundancy analysis (dbRDA) to analyze microbiota between-person diversity (β-diversity), with explicit adjustment for geographical modification.
b. In a sub-sample with metabolites data, use multivariable-adjusted linear regression to analyze individual microbiota, and dbRDA to analyze the overall metabolites, with explicit adjustment for geographical modification. Additionally, use random forest regression to assess the accuracies of

host factors data, microbiota data, and metabolites data in predicting sodium and potassium consumption.

We hypothesized that sodium and potassium are associated with microbiota and metabolites that have been previously linked to cardiovascular disease risk factors.

Aim 2. Examine the association between gut microbiota (16S rRNA) and plasma metabolites with systolic and diastolic blood pressure (SBP & DBP).

a. In the microbiota analysis sample, use multivariable-adjusted linear regression to analyze microbiota within-person diversity (α-diversity) and specific taxa, and permutational multivariate analysis of variance (PERMANOVA) to analyze microbiota between-person diversity (β-diversity).
b. In the metabolomics analysis sample, use multivariable-adjusted linear regression to analyze individual microbiota, which paired with a pathway analysis to identify metabolic pathways associated with SBP and DBP. Additionally, using principal component analysis (PCA) to identify patterns of metabolites associated with SBP and DBP.
c. In a sub-sample with both microbiota and metabolites data, use random forest regression to assess the accuracies of host factors data, microbiota data, and metabolites data in predicting SBP and DBP.

We hypothesized that microbiota and metabolites that have been identified to regulate blood pressure in animal models are associated with SBP and/or DBP, such as short-chain fatty acids (SCFAs).

Aim 3. Examine the association between plasma SCFAs with two adiposity measures, BMI and waist-to-height ratio (WHtR).

a. Use multivariable-adjusted linear regression to investigate the association between the relative abundance of each plasma SCFAs and the total plasma SCFAs with BMI and WHtR.

b. To test whether subjects with higher body mass and/or abdominal adiposity had higher SCFAs was due to higher capacity to ferment carbohydrate, a dietary precursor of SCFAs, we assess whether BMI or WHtR levels modify the associations between carbohydrate, fiber, and high-fiber foods with SCFAs, using multivariable-adjusted linear regression.
c. In a sub-sample with gut metagenome data, use multivariable adjusted linear regression to investigate the associations between gut microbial SCFA producers with BMI and WHtR.

We hypothesized that (1) plasma SCFAs were positively associated with BMI and WHtR; (2) when consuming the same amounts of dietary precursors or SCFAs, adults who had higher BMI and/or higher WHtR had higher plasma SCFAs; (3) microbial SCFA producers were positively associated with BMI and WHtR.

CHAPTER 2. LITERATURE REVIEW

Main cardiovascular disease risk factors

Cardiovascular diseases (CVD) are the leading causes of death around the world and reduce quality of life [1]. From 2006 to 2016, the number of deaths attributed to CVD had increased by 14.5% globally [2]. In contrast, the age-adjusted rate of CVD deaths had decreased by 14.5%, yet this decline has been decelerated substantially probably due to the obesity epidemic [2].

Overweight and obesity are major risk factors for CVD [3,4]. Obesity was positively associated with lifetime risk of CVD, prevalence of hypertension, type 2 diabetes, and dyslipidemia [4], leading to higher medical cost for obesity than normal-weight [5]. In particular, body mass index (BMI) in adolescents was strongly associated with elevated CVD death in adulthood [3]. The worldwide prevalence of overweight and obesity increased by 27.5% in adults from 1980 and 2013 [6]. In the US, the prevalence of obesity in adults has increased significantly in the past decade [2] and was more than doubled since 1970s [7]. As a result, the projected increase in cardiovascular health in the United States (US) due to recent improvement of health behaviors like smoking would be offset by the increased prevalence of obesity [8]. What make it worse, waist circumference as a better marker than BMI in predicting CVD risk [9] increased even greater than expected based on the increase in BMI in the US, Mexico, England and China, especially in young women [10].

High blood pressure is another major risk factor for CVD and mortality [2]. In the US, it was estimated that CVD mortality could be reduced by 30.4% and 38.0% among males and females, respectively, if high blood pressure was eliminated [11]. In fact, the elimination of high blood pressure had the largest reduction on CVD death than all other risk factors in females and all other risk factors except smoking in males [11]. High blood pressure is also a single leading global burden of disease risk

factor [12]. Particularly, China has the greatest absolute burden for elevated blood pressure around the world by 2011 [13], which is projected to continue to expand in the next decade [14]. The worldwide prevalence of high blood pressures has increased substantially over the past two decades [15,16]. Specifically, the age-standardized prevalence of high blood pressure increased by 2.5% from 2000 to 2010, yet the proportion of awareness, treatment, and control remained low, especially in low and middle-income countries (e.g., only 6.9% for control) [16].

Key dietary risk factors for CVD

The key dietary risk factors for elevated adiposity and blood pressure include high sodium, low potassium, and low fiber intakes. The mean level of global sodium consumption in 2010 was estimated to be 3.95 g/day [17], which was well above the dietary recommendation of 2.0 g/day by the World Health Organization (WHO) [18]. Globally, annual cardiovascular deaths attributable to high sodium intake (> 2.0 g) accounted for nearly 10% of deaths from cardiovascular cause [17]. Each additional 1000 mg sodium in 24-h urine excretion (a marker for sodium consumption) was associated with 4.6 and 2.3 mmHg higher systolic and diastolic blood press (SBP and DBP), respectively, whereas each additional 1000 mg potassium in 24-h urine excretion was associated with 3.7 mmHg lower SBP [19]. Sodium intake also show positive associations with body mass and abdominal obesity independent of energy intake [20]. Moreover, high-fiber foods such as whole grains, fruits and nuts were inversely associated with risk of high blood pressure, while red meat, processed meat and sugar sweetened beverage were positively associated with risk for high blood pressure, according to a dose-response meta-analysis of 28 prospective studies [21]. There is also evidence for modest negative associations between consumption of cereal fiber, whole grains, and bran with risk of obesity and CVD [22]. Despite the well-understood benefits of dietary fiber, the meat intakes of fiber and whole grain remained below recommendation in the US [23].

Nevertheless, individual response to these dietary factors may vary considerably, especially for sodium [24]. While some people have little blood pressure changes, others have blood pressure changes

paralleling with dietary sodium intake - a phenomenon known as salt sensitivity [25]. Salt sensitivity is a strong risk factor for cardiovascular disease and death independent of blood pressure [26,27], yet how factors like host genetics and gut microbiota determining individual susceptibility to salt sensitivity is far from being completely explored.

Role of gut microbiota and host metabolites in main CVD risk factors

The recent advancement on high-throughput sequencing has revealed that the gut microbiota are potential determinants of cardiovascular health. For example, intestinal bacteria in glucocorticoid metabolism have been implicated in the development of elevated blood pressure [28-31], such as *Eggerthella lenta* [32], which is involved in the 21-dehydroxylation of cortisol to 21-deoxy-5α-tetrahydrocortisol and 21-deoxy-tetrahydrocortisol [28]. These glycerrhetinic acid like factors (GALFs) can raise blood pressure by inhibiting 11β-hydroxysteroid dehydrogenase-2 [28,33], an important enzyme that prevents water and sodium retention [28,34]. Additionally, numerous experimental studies using animal models have demonstrated the impact of gut microbiota on obesity and high blood pressure [35-37]. For instance, mice colonized with fecal microbiota from the co-twins with obesity had larger increases in adiposity and body weight compared to mice colonized with microbiota from the lean co-twins [35]. Fecal transplantation from spontaneously hypertensive stroke prone rats to Wistar Kyoto rats was associated with greater increase in SBP than Wistar Kyoto rats in the control group, confirming that the elevated blood pressure phenotype is transferable through gut microbiota [36].

Human studies have also indicated that the gut microbiota is altered in obesity and high blood pressure, such as reduced microbial diversity [38-43]. Several fecal microbiota were different between adults with elevated versus normal blood pressure [39-41], including over-expressed *Prevotella, Klebsiella, and Actinomyces* [39,40], which are linked to infections [44-46], and under-expressed *Roseburia spp.* and *Faecalibacterium prausnitzii* [41], which are short-chain fatty acids (SCFAs)-producing bacteria that exerts anti-inflammatory properties [47]. Lyxose and 4-hydroxyhippurate, which are products of gut microbial fermentation [48], were also associated with elevated blood pressure in

Chinese and African American adults [49,50]. However, few observational studies examine gut microbiota and related metabolites together to infer potential biological pathways. In addition, most of these studies compared people with body mass or blood pressure above versus below a single threshold, which may overlook people at borderline. These people are still at risk for CVD, given the dose-response relationship between, BMI, abdominal fatness [51], and blood pressure with CVD mortality [52]. Thus, the exact mechanisms linking gut microbiota to increased adiposity and blood pressure in humans remain unclear.

Moreover, both mice and human studies showed that early exposure to antibiotics may predispose them toward obesity [53-55], whereas probiotics supplementation had anti-obesity and anti-hypertension effects to animal models and humans [37,55-58] through proposed mechanisms including alterations in several specific taxa, reduced expression of pro-inflammatory cytokines, increased fatty acid oxidation, and production of satiety-inducing peptides [55]. Genetic studies added another layer of evidence for the association between gut microbiota, obesity, and blood pressure [31,59-62]. Some host gene loci associated with gut microbiome are in close proximity to gene loci of disease risk factors including elevated blood pressure [31,59,60]. For example, genes in steroid biosynthetic pathway are associated with BP [61] and may shape the gut microbial composition [31,62].

The interplay between diet and gut microbiota in cardiovascular health

The gut microbiota and related metabolites plays a pivotal role in the diet-health relationship. Diet could shape the gut microbiota community [63]. In particular, high sodium diet altered fecal microbiota composition, microbial metabolites, autoimmunity, and protein digestion in murine models [42,64-66], including depleted *Lactobacillus species* and increases in fecal SCFAs and intestinal T helper 17 (T_H17) cells. Concomitant treatment with *Lactobacillus murinus* prevented sodium-induced high blood pressure in mice [42], indicating that gut microbiome could be a potential therapeutic target for sodium-associated diseases. Similarly, in humans, sodium consumption was associated with changes in the microbial composition, circulating microbial-mediated metabolites, and related metabolic pathways

[42,67,68]. For example, high sodium intake was associated with reduced *Lactobacillus* [42], while reduced sodium intake was associated with plasma metabolites from tryptophan and benzoate metabolic pathways mediated by microbiota [67], such as increased 4-ethylphenylsulfate, which has been shown to be positively associated with percent lean body mass in young adults [69]. Another key dietary factors for CVD risk, dietary fiber, is positively correlated with microbiota diversity and polysaccharide-utilizing microbiota, but negatively correlated with protein fermentation products, *Bacteroides*, and *Clostridia* [70]. A mouse model has shown that fermentable fiber could prevent high-fat diet-induced depletion in microbiota, colon atrophy, and metabolic syndrome [71]. Altogether, these studies indicate that diet may affect health through changes in gut microbiota, which could be reflected by circulating metabolites, and modification of gut microbiota may ameliorate or even reverse the harmful effects of diet.

Another pathway of how the gut microbiota contribute to diet-health relationship is through harvesting of dietary energy [72-74]. Conventionally raised mice consumed 28% less calories but had 40% more body fat than germ-free mice [72]. In human adults, SCFAs (e.g., acetate, butyrate, and propionate) produced by gut microbial fermentation of carbohydrate contribute to nearly 10% of extra daily calorie [70]. Both fiber-rich diet and Mediterranean diet have been shown to be associated with increased levels of circulating and fecal SCFAs [75,76], indicating that SCFAs may serve as the link between diet, gut microbiota, and host health. While acetate and propionate are components of citric acid cycle and gluconeogenesis [77], respectively, butyrate may prevent inflammation and carcinogenesis of the colon epithelium [70]. SCFAs have also been shown to improve intestinal mineral absorption and permeability [77,78]. They can interact with G protein-coupled receptors (GPCRs) like GPR41 and GPR43 to affect adipocytes and peripheral organs, such as kidney and pancreas, thus regulating blood pressure, fat oxidation, and energy metabolism [74,79-81]. For example, Gpr41 enhances the blood pressure-reducing effects of propionate [80]. Although several studies have shown that SCFAs were associated with reduced inflammation, improved insulin sensitivity, increased satiety, and increased energy expenditure in mouse model [71,77,82], as well as in human trials of dietary supplementations [83-85], other studies have suggested that SCFAs are indeed sources of extra energy and may promote the

development of obesity [72,73,86,87]. For example, a mouse study showed that SCFAs production outweighed the benefits of soluble fiber in the context of diet-induced obesity by increasing digested energy [87], and a few epidemiology studies of Western populations showed that fecal SCFAs were positively associated with the body mass, total body fat and central adiposity [78,88-91]. Possible reasons for these incongruent results include different population, small sample sizes, and the differences in amounts and bioactivities between the experimentally administered SCFAs supplements and the SCFAs produced by the gut microbiota. However, few population-based studies have investigated the associations between circulating SCFAs and adiposity measures. Circulating SCFAs may better reflect the absorption and the energy-contributing fraction of SCFAs than fecal SCFAs [92].

Current gaps in research and needed work

Prior studies have shown the complex relationships across diet, gut microbiota, microbial metabolites, and main CVD risk factors like obesity and high blood pressure. However, a few major gaps in research remain unexplored. These include the (1) lack of large population-based study with large variations in dietary intakes and CVD risk; (2) lack of studies that couple metabolomics analysis with gut microbiota analysis to reveal potential biological mechanisms, as microbiota-mediated metabolites derived from dietary components are efficient measures of microbial function [93]; and (3) lack of studies in Asians, who are different from other populations in genetics, gut microbiota [92], and dietary habits [94]. For example, Asians have higher sodium intake and different sodium sources than Whites, Hispanics/Latinos, and Blacks [94]. As such, the associations between diet, microbiota, and CVD risk factors observed in Western populations may not be generalizable to Asians.

Specifically, while there is strong evidence supporting the roles of microbiota and metabolites in the deleterious effects of excess dietary sodium on cardiovascular health [42,64-66], few studies have explored the associations between potassium, another key dietary protectors for CVD [19], with microbiota and metabolites in humans. Additionally, many gut microbiome studies of blood pressure in humans were not population-based [39-41], might have overlooked people at borderline of high blood

pressure [39,40], and were unable to fully eliminate the effects of hypertension treatment [41,43]. Given that Chinese adults have habitual high sodium and low potassium intakes [95,96] and low diagnosis and treatment rates for high blood pressure [97], which ensures a natural history of untreated blood pressure and low confounding from treatment effect by design, they are an ideal population for research on dietary sodium and potassium and blood pressure.

Therefore, we used a large well-characterized, population-based adult cohort from the China Health and Nutrition Survey (CHNS), which provides diverse sample from a range of urban and rural communities across 12 provinces and three megacities during rapid urbanization, in order to investigate the following objectives: (1) associations between dietary sodium and potassium consumption with gut microbiota, and circulating metabolites; (2) associations between gut microbiota and circulating metabolites with SBP and DBP; and (3) the associations between circulating SCFAs with adiposity measures and the potential involvement of diet and gut microbiota in these associations. In addition, we incorporated gut metagenome data that provides more detailed taxonomic classification than 16S rRNA data and used methods like multivariate analysis and random forest regression, a machine learning algorithm, to reduce dimensionality and to account for the intricate correlations among gut microbiota and metabolites.

In summary, by comprehensively exploring the relationships between diet, gut microbiota, circulating metabolites, and main CVD risk factors (i.e., blood pressure, body mass, and abdominal adiposity) in a unique Chinese adult cohort, we may improve our current understanding of individual response to diet and the underlying mechanisms of CVD development, thereby shedding light on potential biomarkers and informing future intervention and treatment programs targeting the gut microbiota and host metabolites.

CHAPTER 3. METHODS

Study population

The CHNS is a longitudinal household-based survey designed to capture urbanization and diverse health data from nine provinces in China since 1989, with 2-4 years intervals between each survey round [98]. Counties and cities stratified by income and urbanicity were selected using a stratified probability sample based on multistage, random cluster design. Then, communities and households were randomly selected from these strata that initially represent the national age, gender, and education profiles [99]. Three megacities and three additional provinces were included in 2011 and 2015, respectively, using the same sampling strategy, giving a total number of more than 30,000 participants. By 2015, a total of 12 provinces ((Heilongjiang, Liaoning, Shaanxi, Henan, Hubei, Jiangsu, Shandong, Zhejiang, Guangxi, Guizhou, Hunan, Yunnan) and three megacities (Beijing, Shanghai, Chongqing) that varied substantially in geography, economic development, public resources, and health indicators were surveyed. Each exam round collected thorough community, household, and individual data by trained examiners in high detail, including household composition, occupation, income, education, anthropometry, and health behaviors. The study met the standards for the ethical treatment of participants and was approved by the Institutional Review Boards of the University of North Carolina at Chapel Hill and the National Institute for Nutrition and Health, Chinese Center for Disease Control and Prevention. Informed consents were obtained for all participants.

Dietary assessment

Diet data were collected by trained interviewers using household food inventories and three consecutive validated 24-h diet recalls during three-day home visits randomly from Monday to Sunday to

ensure a mix of weekdays and weekends. These dietary measures not only assessed cooking methods and all foods consumed within the household, but also foods consumed at restaurants and other locations away from home. All household foods and condiments were measured using digital kitchen scales (graduation: 1g) and allocated to each member based on the proportions consumed as reported in the individual 24-h diet recalls. To estimate the average daily intakes of nutrients across three days, individual-level diet data were linked to a Chinese food composition table (FCT) that includes more than 2500 foods [95]. For a few imported foods, nutrients contents were estimated using Taiwan, Hong Kong, Japan, or the USDA FCTs [95]. The protocol for total energy measurement was validated by doubly labeled water (Pearson correlation coefficient, males: 0.56, females: 0.60) [100]. Specifically, insoluble fiber was measured by the neutral detergent method, and for sodium and potassium, we calculated all measured foods and condiments, including added salt and soy sauce in mixed dishes. Our measurement tools for sodium and potassium intakes were validated by three-consecutive 24-hour urine samples (Pearson correlation coefficients, sodium: 0.58, potassium: 0.59) in an independent sample from one of our survey provinces, using para-aminobenzoic acid as a marker for completion of 24-hour urine collection [95].

Gut microbiota data collection and processing

Stool samples were collected at home in 2015 from two sub-cohorts of participants aged 18-80 years, who had been trained to using the QIAGEN collection kit (QIAGEN, Hilden, Germany) following standardized procedures from a modified Human Microbiome Project (HMP) protocol. These two cohorts are the China Nutritional Transition Cohort Study (CNTCS, n=2,164) across all 12 provinces and three megacities, and the China Microbiome Study (n=1,226) across four southern provinces: Henan, Guizhou, Hunan, and Guangxi. The total number of participants in these two cohorts was 3,208, as 182 adults were included in both cohorts.

Stool samples were frozen at -20°C immediately after collection and were sequenced by Novogene Bioinformatics Technology Co., Ltd., Tianjin, China in random order. Bacterial DNA was

extracted using TIANGEN DNA extraction kits (TIANGEN Biotech, Beijing, China) and 16S rRNA sequencing targeting the V4 hypervariable region was performed using primers 515F/806R on the Illumina MiSeq PE250 platform. The sequencing generated 12,528-7,7104 sequences in CNTCS and 21,648-89,427 sequences in China Microbiome Study. The raw sequencing reads were processed using the QIIME pipeline [101], with forward and reverse reads merged with fastq-join, and filtered using a minimum quality score of 20. Operational Taxonomic Units (OTUs) were identified using open-reference method based on a threshold of 0.97. Chimeric OTUs were detected by ChimeraSlayer [102] and removed. Taxonomy was assigned based on the SILVA databases (Release 128). No sample was filtered out due to low quality. A total of 1,472 genera were detected. In addition, the whole metagenome was measured in 214 adults aged 30-68 years from the Hunan and Guizhou, the majority of whom were included in the China Microbiome Study (79%). After filtering human DNA from the sequencing reads, we annotated the reads with MetaPhlAn2 based on the ChocoPhlAn database for taxonomic composition [103]. To correct for different sequencing depth across samples, we normalized and $\log_{10}$ transformed the raw taxonomic counts in the 16S rRNA data and the metagenome data as follows [104]:

$$log_{10}\left(\frac{taxa\ j\ count\ for\ sample\ i}{total\ taxa\ count\ in\ sample\ i}\ \text{x}\ average\ number\ of\ taxa\ count\ per\ sample + 1\right).$$

Plasma metabolomics profiling

Fasting blood samples were collected within 3-days of fecal sample collection using venipuncture with EDTA as an anticoagulant, centrifugated to prepare plasma, and stored at -80°C until analyzed. All sites followed the same protocol for the collection, processing, and storage. In 500 adults aged 30-68 years from Hunan and Guizhou, the majority of whom were included in the China Microbiome Study (87%), non-targeted metabolomics analysis was performed using Metabolon platform (Durham, NC) consisting of Waters Acquity ultrahigh-performance liquid chromatography coupled to a Thermo Scientific Q-Exactive high-resolution Mass spectrometry at Metabolon's partner campus in China [105]. More detailed information on the Metabolon platform has been described elsewhere [106]. Briefly,

plasma samples were extracted using methanol solvent and analyzed with several types of controls, including extracted water samples as process blanks and pooled experimental samples as technical replicate. Signals in the metabolomics data were extracted and peak identified, with quality control processed using Metabolon's software and hardware. Chemicals were identified and differentiated by matching to the mass-to-charge ratio, rendition time/index, and chromatographic data in the Metabolon reference library of authenticated standards, which was created by acquiring data for more than 3,300 purified standard compounds analyzed under the same conditions as the study samples. The metabolomics analysis resulted in the detection of 1,108 compounds in our sample. Metabolon rescaled the raw area under the peaks of each metabolite within the same run-day to a median of 1 (i.e. median-normalization) to correct for differences in instrument inter-day tuning, with values below detection limits imputed by the minimum. We $\log_2$ transformed the metabolites relative abundance to ensure normality.

Anthropometry and blood pressure

Anthropometry data were collected during physical examination by trained examiners. Weight was measured to the nearest 0.1 kg in light clothing using calibrated beam scales. Height was measured without shoes to the nearest 0.1 cm using portable stadiometers. Waist circumference was measured to the nearest 0.1 cm at midway between the lowest rib and iliac crest using non-elastic tape. We calculated BMI as weight divided by squared height (kg/m^2) and WHtR as waist circumference divided by height. We defined high BMI as BMI $\geq$24 kg/m^2 according to the Chinese cut-point [107] and high WHtR as WHtR $\geq$0.5, which has been determined to be the optimal cut-off point to indicate cardiovascular diseases risk in Chinese adults [108]. Resting blood pressure was measured by experienced physicians, who had completed a 7-day training session and passed a comprehensive test for the reliability of measurements. SBP and DBP were measured in triplicates using a standard mercury sphygmomanometer on the right arm after a 10 min seated rest, with a 30-second interval between cuff inflation. We used the average of the three readings as our measures of SBP and DBP.

Sociodemographic and behavioral factors

Sociodemographic and behavioral data at community-level (i.e., urbanization), household-level (i.e., household income), and individual-level (i.e., age, sex, province/megacity) were collected using interviewer-administered questionnaires. Specifically, we assessed urbanization using the urbanization index, a validated index that encompasses 12 dimensions of urbanization such as population density, health infrastructure, sanitation, and transportation [109]. We grouped urbanization index into tertiels to represent low, middle, and high levels of urbanization. We calculated per capita household income by dividing the household income by the number of household members, and then categorize it into tertiles to represent low, middle, and high levels of income. We dichotomized educational attainment by the completion of high school education. We grouped occupation into six five categories: not-working, agriculture workers (e.g., farmer, fisherman, hunter), laborers (e.g., craftsman, logger), professional (e.g., doctor, teacher), and manager (e.g., government official, director) and other (e.g., athlete, artist). We dichotomized educational attainment by high school completion. We measured total physical activity in METs/week using seven-day recalls of occupational, transportation, domestic, and leisure activities. We categorized physical activity by tertiles to indicate low, middle, and high levels of physical activity. We dichotomized fried food intake by any/no consumption and calculated percent calories (%kcal) from animal-source foods by dividing energy intake from animal-source foods by the total energy intake. We defined smokers as individuals who ever smoked cigarettes across all completed surveys and alcohol consumers as individuals who drank alcohol during the past year.

CHAPTER 4. ASSOCIATIONS OF SODIUM AND POTASSIUM CONSUMPTION WITH GUT MICROBIOTA AND HOST METABOLITES IN A POPULATION-BASED STUDY OF CHINESE ADULTS

Overview

There is increasing evidence that sodium consumption alters gut microbiota and host metabolome in murine models and small studies in humans. However, there is a lack of population-based studies that capture large variations in sodium consumption as well as potassium consumption. Thus, we examined the associations of energy-adjusted dietary sodium (mg/kcal), potassium, and sodium-to-potassium (Na/K) ratio with microbiota and plasma metabolome in a well-characterized Chinese cohort with habitual excessive sodium and deficient potassium consumption.

We estimated dietary intakes from validated three-consecutive 24-h recalls and household inventories. In 2833 adults (18-80 years old, 51.2% females), we analyzed microbial (genus-level 16S rRNA) between-person diversity, using distance-based redundancy analysis (dbRDA), and within-person diversity and taxa abundance using linear regression, accounting for geographic variation in both. In a sub-sample (n=392), we analyzed the overall metabolome (dbRDA) and individual metabolites (linear regression). P-values for specific taxa and metabolites were false discovery rate-adjusted (q-value).

We found that sodium, potassium, and Na/K ratio were associated with microbial between-person diversity (dbRDA p-value<0.01) and several specific taxa with large geographic variation, including pathogenic Staphylococcus and Moraxellaceae, and short-chain fatty acids (SCFA)-producing Phascolarctobacterium and Lachnospiraceae (q-value<0.05). For example, sodium and Na/K ratio were positively associated with Staphylococcus and Moraxellaceae in Liaoning, while potassium was positively associated with two genera from Lachnospiraceae in Shanghai. Additionally, sodium, potassium, and Na/K ratio were associated with the overall metabolome (dbRDA p-value≤0.01) and

several individual metabolites, including butyrate/isobutyrate and gut-derived phenolics like 1,2,3-benzenetriol sulfate, which was negative associated with sodium in Guizhou (q-value<0.05).

In conclusion, our findings suggest that sodium and potassium consumption is associated with taxa and metabolites that have been implicated in cardiometabolic health, providing insights into the potential roles of gut microbiota and host metabolites in the pathogenesis of sodium- and potassium-associated diseases. More studies are needed to confirm our results.

Introduction

Excessive dietary sodium intake and inadequate dietary potassium intake contribute to hypertension and cardiovascular disease (CVD) [95,110,111], through mechanisms involving the renin-angiotensin-aldosterone system and oxidative stress [112,113]. Recent advances in high-throughput sequencing have revealed that intestinal microbes are dependent on diet and may have fundamental impacts on host metabolome and physiology, including blood pressure regulation [42,114]. Therefore, elucidating the relationships between key diet risk factors, such as sodium and potassium, with the gut microbiota and circulating metabolites is essential in understanding the roles of microbiota and related metabolites in diet-associated diseases.

Evidence from murine models suggests that a high sodium diet changes fecal microbiota composition and function, with depletion of *Lactobacillus* and increases in fecal short chain fatty acids (SCFA) and microbial-dependent intestinal T helper 17 (T_H17) cells [42,64-66]. Concomitant treatment with *Lactobacillus murinus* prevented sodium-induced hypertension in mice [42], indicating that gut microbiome could be a potential therapeutic target for sodium-associated diseases. In addition, metabolomics studies have revealed potential pathways underlying the sodium-health relationships. In 119 US adults from the cross-over sodium intake feeding trail within the Dietary Approaches to Stop Hypertension (DASH)-Sodium trial, reduced sodium intake was associated with increased plasma metabolites from the microbiota-mediated tryptophan and benzoate metabolic pathways [67], such as 4-ethylphenylsulfate that has been linked to lean body mass in adults [69]. In a double-blinded, cross-over

trial of 64 untreated UK patients with hypertension, sodium reduction was associated with elevated serum methionine sulfone and β-hydroxyisovalerate, which were associated with reduced diastolic and systolic blood pressure in the same sample, respectively [68]. However, there is a lack of population-based studies that capture large variations in sodium consumption with paired microbiome and metabolomics data for a more comprehensive investigation of these complex relationships. There has also been a lack of studies examining dietary potassium in Asians, who have higher sodium intake and different sodium sources than Whites, Hispanics/Latinos, and Blacks [94].

To address the above knowledge gaps, we used data from a population-based cohort of Chinese adults with habitual high sodium and low potassium intakes [95,96] to study two questions. First, we examined the association between sodium and potassium consumption and gut microbiota in 2,833 adults from 12 provinces and three megacities. Second, to further understand potential biological responses to sodium and potassium, we examined the association between sodium and potassium and plasma metabolites in a sub-sample of 392 adults from two southern provinces.

Methods

Study sample

We used data from the China Health and Nutrition Survey (CHNS) collected in 2015 during fall (primarily) and winter. Eligible participants were adults aged 18-80 years who had gut microbiome and diet data (n=3,156, Figure S4.1). Participants were excluded if they used antibiotics within the past 6 months, ever had inflammatory bowel disease, irritable bowel syndrome, or bowel removal, or currently had diarrhea (n=217). Participants were further excluded if they were pregnant (n=1) or had extreme energy intake (<500 kcal, n=8), implausible sodium consumption (>10 mg/kcal, n=1), or missing covariates (n=96), resulting in an microbiota analysis sample of 2,833 adults, among whom 905 from the China Microbiome Study and 1,928 from the China Nutritional Transition Cohort Study (CNTCS), which are two sub-cohorts of the CHNS. A subset of 392 adults living in adjacent provinces (Hunan and Guizhou) were included in the metabolomics analysis sample.

Measures

We standardized sodium and potassium by total energy intake in unit of mg/kcal to account for over- and under-reporting and the correlation with energy intake, and divided sodium by potassium to calculate sodium to potassium (Na/K) ratio. We defined excessive sodium (≥2 g) and deficient potassium (<3.5 g) consumption according to the WHO [115]. Pro/prebiotics intake in past month, non-steroid anti-inflammatory drugs (NSAIDs) intake in past two weeks, and proton pump inhibitors (PPIs) intake in past month were assessed by questionnaires admistered before fecal sample collection.

Statistical analysis

In descriptive analysis, we compared continuous variables and categorical variables across provinces and megacities using analysis of variance (ANOVA) and Chi-squared test, respectively.

Primary outcomes were gut microbial measures at genus level. We first examined the associations of sodium density, potassium density, and Na/K ratio with microbial diversity measures using R package vegan [116]. For within-person diversity (α-diversity) assessed by Shannon index and richness (number of distinct genera per subject), we used a linear regression. For between-person diversity (β-diversity) assessed by principal coordinates analysis (PCoA), we used distance-based redundancy analysis (dbRDA) [117] based on Bray-Curtis distance, a multivariate analysis that did not provide the direction of associations, followed by an analysis of variance (ANOVA) test with 999 permutations to estimate p-value. We then used a linear regression to assess the associations of sodium density, potassium density, and Na/K ratio with 159 specific taxa, after filtering rare taxa presented in less than 25% of participants to account for spurious findings. All analyses were adjusted for the following potential confounders based on *a priori* knowledge: age, sex, provinces or megacities [95,118], batch or plate runs, urbanization [119], occupation, education, income, total energy intake, %kcal from animal-source foods (the strongest indicator of Westernized diet in China) [120], fried food intake, physical activity [121], smoking [122], alcohol, pro/prebiotics, non-steroid anti-inflammatory drugs (NSAIDs), and proton pump inhibitors

(PPIs) intakes [123]. In addition, we conducted secondary analysis at OTU level, using linear regression for individual OTUs (n=256 after excluding rare OTUs), to contribute to understanding variation in genus-level results.

Given that another study has found large geographic variation (provinces and megacities) in the CHNS gut microbiota sample [118], we accounted for this geographic variation by including province/megacity (categorical variable) and an interaction term of sodium density, potassium density, or Na/K ratio with province/megacity in the model. In addition to a test for interaction, we assessed the overall association using a joint analysis that simultaneously tests the main effect (sodium density, potassium density, or Na/K ratio) and interaction term for province/megacity [124]. This joint test is common in genetic studies with potential interactions since it offers more statistical power than other methods in the presence of interaction and comparable statistical power to other methods when there is no interaction [124]. The interaction and joint analyses were examined using Wald test in linear regression and using partial dbRDA conditioned on the rest of the model variables. For α- and β-diversity measures, the interaction term was removed if the p-value of joint test was >0.10.

In the subsample, we repeated the above analyses for metabolomics data (secondary outcome) using dbRDA for the overall metabolome and linear regression for individual metabolites, with adjustment of batch run. We conducted exploratory analysis of the associations between sodium density-, potassium density-, and/or Na/K ratio- associated taxa and metabolites using multivariable-adjusted linear regression. To assess which data had the strongest association with sodium density, potassium density, and Na/K ratio intakes, we compared prediction accuracies of these dietary outcomes by host factor (18 model covariates), microbiota, metabolite data, and all the possible one by one permutations of the combinations of host factors, microbiota, and metabolite data, using random forest regressions (100 trees) [125]. We conducted pairwise comparisons of root mean squared errors (RMSEs) of each model using the 5 iterations of 2-folds cross-validation modified paired t-test, which is powerful to compare the performance of learning algorithms with acceptable Type I error [126].

We conducted statistical analyses in R 3.6.0 (http://www.r-project.org) and Python 3.5.1 (https://www.python.org). All statistical tests were two-sided. For comparisons across all taxa and all metabolites in linear regression, p-values were adjusted using Benjamini-Hochberg method (false discovery rate, q-value) [127] for sodium density, potassium density, and Na/K ratio separately as part of each test of three separate hypotheses for sodium density, potassium density, and Na/K ratio.

Results

Sample characteristics

The microbiota analysis sample had 51.2% females and a mean age of 51.6 years (Table 4.1). Gut microbial α-diversity (Shannon index and richness), physical activity, urbanization, income, education, occupation, and intakes of sodium, potassium, Na/K ratio, energy, animal-source foods, fried food, and pre/probiotics were different across provinces and megacities (p-value<0.001). The megacity Chongqing had highest Shannon index and richness and the lowest potassium intake, whereas Shanghai, a megacity with 77% and 69% participants at high urbanization and income level, respectively, had the highest potassium and the lowest Na/K ratio intake. Yunnan, one of the least urbanized provinces (67% at low urbanization level), had the lowest Shannon index and animal-source food intake, but the highest Na/K ratio intake, while Zhejiang, one of the provinces with the highest income (>67% at high income level), had the highest sodium intake. In the metabolomics analysis sample, Hunan had higher urbanization and intakes of sodium and fried food but lower microbial richness than Guizhou (p-value<0.001, Table S4.1).

Microbiota analysis

First, we evaluated the overall measures of the gut microbiota composition and found that sodium density, potassium density, and Na/K ratio were not associated with α-diversity measures (Table 4.2), but associated with β-diversity, which varied across provinces and megacities (interaction and joint test dbRDA p-value<0.01, Table 4.3). The microbial β-diversity was visualized with multiple dimension

scaling (MDS) in Figure S2-4, which show no clear separation of microbiota by sodium density, potassium density, or Na/K ratio.

Next, we examined specific taxa and found that at joint test q-value<0.10, sodium density was associated with eight taxa, including *Staphylcoccus, Moraxellaceae, Phascolarctobacterium, Salinicoccus,* and *Jeotgalicoccus* (Figure 4.1A); potassium density was associated with 30 taxa, including *Pseudomonas*, *Staphylcoccus*, *Dorea*, *Leuconostocaceae*, and *Oscillospira* (Figure 4.1B); and Na/K ratio was associated with 54 taxa, including *Moraxellaceae, Pseudomonas, Lactobacillales, Staphylcoccus, and Microbacterium* (Figure 4.1C). These associations showed large geographic variations. Province- and megacity-specific model estimates are shown in Table S4.2-4.4, respectively. For example, sodium density was negatively associated with *Moraxellaceae* in Beijing [coefficient (95% confidence interval): -0.10 (-0.17, -0.04)], but positively associated with it in Liaoning [0.14 (0.07, 0.21)] and Shanghai [0.18 (0.09, 0.27)].

Secondary analysis using OTU-level data was consistent with genus-level results. We observed large geographic variation in the following associations at joint test q-value<0.10, sodium density with *Phascolarctobacterium* (Table S4.5); potassium density with six OTUs, including *Dorea, Ruminococcaceae_UCG-014,* and *Weissella* (Table S4.6); and Na/K ratio with 36 OTUs, including *Ruminococcus_2, Lachnospiraceae, Blautia, Phascolarctobacterium, Megamonas, Ruminococcaceae_UCG-014, Catenibacterium,, Coprococcus_2, Clostridium_sensu_stricto_1, Akkermansia, Ruminococcus_1,* and *Prevotella* (Table S4.7).

Metabolomics analysis

We first examined the overall metabolome and found that sodium density, potassium density, and Na/K ratio were associated with the overall metabolome at dbRDA p-value≤0.01 (Table 4.4), which were visualized with MDS in Figure S4.5-4.7. Then, for individual metabolites, we found that at joint test q-value<0.10, sodium density was associated with four metabolites: N6-methyladenosine form purine metabolic pathway and gut-derived phenolics, 1,2,3-benzenetriol sulfate, 3-methoxycatechol sulfate, and

4-methylcatechol sulfate (Figure 4.2A); potassium density was associated with 6-oxopiperidine-2-carboxylate from lysine metabolic pathway (Figure 4.2B); and Na/K ratio was associated with 15 metabolites, including a fibrinogen cleavage peptide (DSGEGDFXAEGGGVR), N6-methyladenosine, thyroxine, two eicosanoid (5-HETrE, 5-HETE), and the microbiota-mediated SCFAs, butyrate/isobutyrate and isovalerate (also a branched-chain amino acid intermediary) (Figure 4.2C). We show province-specific estimates in Table S4.8, respectively.

Integrated analysis of microbiota and metabolite data

We examined the associations between sodium density-, potassium density-, and/or Na/K ratio-associated taxa (n=67) and metabolites (n=18) and found that *Coriobacteriaceae* and *Ruminococcaceae* were positively associated with 4-methylcatechol sulfate (q-value<0.10, Table S4.9). In random forest analysis assessing microbiota or metabolite data as a whole, we found that metabolite data and the combinations of metabolite + host factor data, microbiota + metabolite data, and microbiota + metabolite + host factor data had higher accuracy (lower RMSE, q-value<0.05) than microbiota data in predicting sodium density and Na/K ratio (Figure 4.3). Metabolite data and host factor data had comparable accuracies in predicting sodium density, potassium density, and Na/K ratio, and adding microbiota data to the combination of metabolite + host factor data made no difference in prediction accuracy of these three diet outcomes.

Discussion

In this study, we investigated the association of dietary sodium and potassium consumption with gut microbiota and host metabolites in a population-based cohort of Chinese adults with habitual excessive sodium intake and deficient potassium intake [95,96]. We found that independent of a wide range of sociodemographic and behavioral factors and after accounting for geographic variations, energy-adjusted sodium (i.e., density in mg/kcal), potassium, and Na/K ratio were associated with the microbial between-person diversity (β-diversity) and several microbial groups, including infectious pathogens, like

Staphylococcus [128] and *Pseudomonas* [129], and taxa that have been linked to CVD risk factors, like *Dorea* [130], *Ruminococcus*, *Ruminococcaceae* [131], and *Lachnospiraceae* [132]. In sub-sample analysis, we found that dietary sodium, potassium, and Na/K ratio were associated with the overall metabolome and several metabolites that involved in inflammation and etiology of CVD, including three gut-derived phenolics (1,2,3-benzenetriol sulfate, 3-methoxycatechol sulfate, and 4-methylcatechol sulfate) [133] and two SCFAs (butyrate/isobutyrate and isovalerate) [134]. These results suggest that gut microbiota and related-metabolites may play important roles in sodium- and potassium-associated diseases.

Our findings add evidence to the sodium-microbiota associations in a large, free-living human population and were consistent with previous studies [42,64-66]. High sodium diet has been found to alter the gut microbiota composition and function in murine models, as reflected by decreases in *Lactobacillus* and increases in *Lachnospiraceae*, *Ruminococcus*, and fecal SCFA levels [42,64-66]. Moreover, there is little known about microbiota associated with potassium, another well-established dietary risk factor for CVD [95,110,111]. The sodium- and potassium-associated taxa we found have been implicated in CVD risk. For example, *Dorea* and *Ruminococcus* were positively associated with body mass index in Swedish adults [130]; *Lachnospiraceae* and *Ruminococcaceae* were related to lower long-term weight gain in females from TwinsUK [131]; *Lachnospiraceae* and *Blautia* were correlated with metabolic impairment in Austria older adults [132]; and *Eggerthella* and *Prevotella* were associated with hypertension in Chinese adults [39,41]. Furthermore, we found that in Liaoning, Henan and Shanghai, sodium consumption was positively associated with pathogenic bacteria including *Staphylococcus*, which causes a wide variety of severe infections [128], and *Moraxellaceae*, a biomarker for Crohn's disease [135], indicating that high sodium intake may increase the susceptibility to gut infection and inflammation. Indeed, sodium exposure has been shown to enhance pro-inflammatory cytokine production in human intestinal mononuclear cells and high-sodium diet exacerbated colitis in mice [136].

A probable mechanism of how gut microbiota influence host physiology is through SCFAs, such as acetate, butyrate, and propionate, which are primarily produced from bacterial fermentation of

carbohydrates [134]. Animal studies suggest that SCFAs are potentially beneficial to cardiometabolic health by modulating gut barrier function, immunomodulation, glucose homeostasis, and blood pressure [80,134]. However, human observational studies suggest that higher levels of fecal and plasma SCFAs associate with poorer gut health and higher obesity and hypertension risk [73,78]. We found that Na/K ratio consumption was positively associated with isovalerate and butyrate/isobutyrate in Guizhou. Similarly, the DASH cross-over sodium intake feeding trial showed reduced plasma isovalerate after a month-long sodium-restricted diet [67]. In addition, we found that in Guizhou province, sodium consumption was negatively associated with 1,2,3-benzenetriol sulfate, 3-methoxycatechol sulfate, and 4-methylcatechol sulfate, which are phenolics derived from microbiota conversion of dietary polyphenols that potentially have anti-inflammatory bioactivity [133]. These results are consistent with studies showing high sodium induced inflammation in mice [42,136]. Our integrated analysis of microbiota and metabolites showed that the Na/K ratio-associated *Coriobacteriaceae* and *Ruminococcaceae* were positively associated with 4-methylcatechol sulfate. *Coriobacteriaceae* has been shown to be involved in phenolic conversion [137]. However, we may lack statistical power to detect associations for other microbiota-mediated metabolites, like SCFAs, in the sub-sample. Larger population samples with diet, microbiota, and metabolite data are needed to allow more complex integrated analysis.

The strengths of our study include the well-characterized, population-based cohort with large variations in sodium and potassium consumption. Host factors collected from standardized and validated instruments allowed us to control for a wide range of potential confounders. Our large and diverse cohort allowed us to examine potential effect modification by geographic locations, which explained the largest variation (17.9%) in gut microbiota in our sample [118], compared to all other host factors, including age and sex (<1%). We observed that the associations between sodium and potassium consumption with microbiota varied substantially across provinces and megacities, but not by age or sex, indicating that geographical variation should be considered in future microbiota analyses. We also observed geographic variation using OTU-level data with more detailed taxa identification than genus-level data. The large geographic variation may relate to basal differences in gut microbiota across provinces and megacities

[118], as well as different dietary sources of sodium and potassium [95]. Our analysis is a preliminary step to identify associations across provinces and megacities. Further province- and megacity-specific analyses are needed to delineate mechanisms underlying the observed geographic variation.

A limitation of our study is the potential measurement errors in diet assessment tools. Although 24-h dietary recall has shown poorer performance in estimating sodium intake in Chinese adults than 24-h urine [138], our estimation of sodium and potassium consumption was based on both three-consecutive 24-h recalls and household inventories, and had been validated by 24-h urine [95]. Given that a majority of our unique sample consumed a high-sodium and low-potassium diet, we acknowledge that our findings have limited generalizability to populations with lower sodium and higher potassium intake, common to less urbanized areas. We were unable to adjust for former smoking due to small numbers of male (n=89) and female (n=34) former smokers. Our analyses were limited to microbial community structures using 16S rRNA data and thus specific pathways of relevant microbial functional genes could not be established. Our study also lacks independent sample for replication and repeated measures of microbiota and metabolites to model changes and test stability. Whereas it has been previously reported in the CHNS that microbiota was stable over two weeks [119], circulating metabolites were dynamic [139].

Conclusion

We provide substantial observational evidence to the associations of sodium and potassium consumption with gut microbiota and plasma metabolites in a population-based cohort of Chinese adults with habitual excessive sodium intake and inadequate potassium intake. In line with murine models and smaller human studies, we show that sodium, potassium, and Na/K ratio consumption is associated with microbiota and metabolites related to inflammation and CVD risk factors. Taken together, our findings suggest the roles of the gut microbiota and related metabolites in the diet-health relationship. More studies are needed to replicate our results and fully elucidate the biological pathways linking dietary sodium and potassium to CVD outcomes.

Tables and Figures

Table 4.1. Characteristics of the gut microbiota analysis sample by provinces and megacities, part 1

	Total	Beijing	Heilongjiang	Liaoning	Shaanxi	Henan	Jiangsu	Shandong	Shanghai
N	2,833	112	206	127	110	325	134	118	130
Shannon index[2]	2.6 (0.3)	2.5 (0.3)	2.6 (0.3)	2.7 (0.3)	2.5 (0.3)	2.5 (0.3)	2.6 (0.3)	2.6 (0.3)	2.5 (0.4)
Richness[3]	94.6	79.3	91.6	91.3	87.5	91.3	86.2	84.4	112.7
	(40.2)	(24.1)	(34.9)	(25)	(21.2)	(37.2)	(14.8)	(34.7)	(66.6)
Age, year	51.6	50.8	51.3	52.2	50.1	52.2	52.5	52.1	50.6
	(12.6)	(12.8)	(12.7)	(14.7)	(14.6)	(11.3)	(14.5)	(13.9)	(13.9)
Females, n (%)	1450	53	111	63	54	181	66	56	65
	(51.2)	(47.3)	(53.9)	(49.6)	(49.1)	(55.7)	(49.3)	(47.5)	(50)
Sodium (Na)[4], mg	4188.1	4163.9	4619.6	4123.2	3948.1	4526.1	4294.8	4274.4	3913.9
	(2176.9)	(2553.5)	(2379.4)	(2267.5)	(2293.9)	(2307.2)	(2102.5)	(2020.5)	(2041.1)
Na density (mg/kcal)[4]	2.4 (1.4)	2.6 (1.9)	2.7 (1.4)	2.6 (1.5)	2.1 (1.4)	2.5 (1.5)	2.3 (1.1)	2.2 (1.1)	2.2 (1.1)
Excessive Na[5], n	2517	97	189	106	92	301	122	103	116
(%)	(88.8)	(86.6)	(91.8)	(83.5)	(83.6)	(92.6)	(91.0)	(87.3)	(89.2)

Potassium (K)[4], mg	1575.5 (682.7)	1633.1 (734.9)	1545.6 (605.9)	1535 (643.1)	1582.7 (765)	1420 (636)	1675.5 (679.2)	1678.1 (707.2)	1884 (856.9)
K density (mg/kcal)[4]	0.8 (0.3)	1 (0.6)	0.9 (0.2)	0.9 (0.3)	0.8 (0.3)	0.7 (0.3)	0.9 (0.3)	0.8 (0.2)	1 (0.4)
Deficient K[5], n (%)	2775 (97.9)	110 (98.2)	203 (98.5)	126 (99.2)	107 (97.3)	319 (98.2)	132 (98.5)	113 (95.8)	123 (94.6)
Na/K ratio[4]	3 (2)	2.8 (1.8)	3.4 (2.3)	3 (1.9)	3 (2.4)	3.7 (2.7)	2.9 (1.5)	2.8 (1.5)	2.3 (1.3)
Energy intake[4], kcal	1906.2 (623.5)	1709.8 (573.3)	1829.3 (631)	1644.4 (566.6)	2104 (724.8)	2040 (682.9)	1941.8 (632.1)	2051.9 (624.7)	1837.8 (533.4)
Animal-source foods[4], %kcal	18.2 (13.4)	12.6 (8.7)	9.3 (8.8)	12.5 (11.3)	25.7 (13.5)	7 (8.9)	16.6 (12.6)	11.9 (9.2)	20.1 (10.8)
Fried food intake[4], n (%)	720 (25.4)	60 (53.6)	62 (30.1)	36 (28.4)	34 (30.9)	82 (25.2)	39 (29.1)	49 (41.5)	55 (42.3)
Ever smoked, n (%)	1109 (39.2)	39 (34.8)	80 (38.8)	52 (40.9)	36 (32.7)	131 (40.3)	54 (40.3)	43 (36.4)	41 (31.5)
Alcohol use, n (%)	850 (30)	38 (33.9)	53 (25.7)	40 (31.5)	45 (40.9)	94 (28.9)	34 (25.4)	41 (34.7)	23 (17.7)
Pre/probiotics use, n (%)	18 (0.6)	0 (0)	0 (0)	0 (0)	0 (0)	11 (3.4)	0 (0)	0 (0)	0 (0)

NSAIDs use, n (%)	35 (1.2)	3 (2.7)	2 (1.0)	1 (0.8)	1 (0.9)	5 (1.5)	0 (0)	2 (1.7)	3 (2.3)
PPIs use, n (%)	18 (0.6)	0 (0)	2 (1.0)	2 (1.6)	0 (0)	2 (0.6)	1 (0.7)	1 (0.8)	1 (0.8)
Physical activity[6], n (%)									
Low	944 (33.3)	42 (37.5)	74 (35.92)	44 (34.65)	24 (21.8)	124 (38.2)	42 (31.3)	46 (39.0)	52 (40)
Middle	949 (33.5)	46 (41.1)	71 (34.5)	42 (33.1)	41 (37.3)	86 (26.5)	43 (32.1)	43 (36.4)	56 (43.1)
High	940 (33.2)	24 (21.4)	61 (29.6)	41 (32.3)	45 (40.9)	115 (35.4)	49 (36.6)	29 (24.6)	22 (16.9)
Urbanization[7], n (%)									
Low	937 (33.1)	0 (0)	109 (52.9)	35 (27.6)	37 (33.6)	175 (53.8)	28 (20.9)	26 (22.0)	0 (0)
Middle	957 (33.8)	45 (40.2)	16 (7.8)	36 (28.4)	52 (47.3)	120 (36.9)	40 (29.9)	83 (70.3)	53 (40.8)
High	939 (33.2)	67 (59.8)	81 (39.3)	56 (44.1)	21 (19.1)	30 (9.2)	66 (49.2)	9 (7.6)	77 (59.2)
Income[8], n (%)									
Low	948 (33.5)	12 (10.7)	50 (24.3)	24 (18.9)	34 (30.9)	171 (52.6)	25 (18.7)	25 (21.2)	6 (4.6)
Middle	943 (33.3)	38 (33.9)	76 (36.9)	40 (31.5)	32 (29.1)	92 (28.3)	43 (32.1)	51 (43.2)	55 (42.3)
High	942 (33.2)	62 (55.4)	80 (38.8)	63 (49.6)	44 (40)	62 (19.1)	66 (49.2)	42 (35.6)	69 (53.1)
High school completion, n (%)	1030 (36.4)	88 (78.6)	80 (38.8)	55 (43.3)	43 (39.1)	85 (26.2)	48 (35.8)	54 (45.8)	90 (69.2)

Occupation[9], n (%)									
Not-working	1410 (49.8)	55 (49.11)	109 (52.9)	51 (40.2)	41 (37.3)	213 (65.5)	62 (46.3)	61 (51.7)	59 (45.4)
Agriculture worker	343 (12.1)	0 (0)	30 (14.6)	19 (15.0)	20 (18.2)	36 (11.1)	4 (3.0)	4 (3.4)	0 (0)
Laborers	673 (23.8)	20 (17.)	24 (11.6)	35 (27.6)	28 (25.5)	50 (15.5)	52 (38.8)	26 (22.0)	37 (28.5)
Professional	203 (7.2)	22 (19.6)	20 (9.7)	11 (8.7)	13 (11.8)	18 (5.5)	5 (3.7)	7 (5.9)	7 (5.4)
Manager	204 (7.2)	15 (13.4)	23 (11.2)	11 (8.7)	8 (7.3)	8 (2.5)	11 (8.2)	20 (17.0)	27 (20.8)
High blood pressure[10], n (%)	1844 (65.1)	70 (62.5)	144 (69.9)	94 (74.0)	65 (59.1)	256 (78.8)	99 (73.9)	86 (72.9)	81 (62.3)
Overweight[10], n (%)	1412 (50)	64 (57.1)	129 (62.6)	78 (61.4)	45 (40.9)	191 (59.3)	72 (53.7)	69 (58.5)	61 (47.3)

Table 4.1. Characteristics of the gut microbiota analysis sample by provinces and megacities, part 2

	Zhejiang	**Chongqing**	**Guangxi**	**Guizhou**	**Hubei**	**Hunan**	**Yunnan**	**p-value**
N	123	125	412	283	117	390	121	
Shannon index[2]	2.5 (0.3)	2.7 (0.3)	2.6 (0.3)	2.6 (0.3)	2.5 (0.3)	2.6 (0.3)	2.4 (0.3)	<0.001
Richness[3]	87.7 (43.7)	174.1 (59.2)	92.8 (27.2)	96.8 (45)	91.2 (25.7)	85.5 (27.7)	90.8 (26.9)	<0.001
Age, year	52 (13.3)	52 (14.7)	51.3 (10.6)	50.8 (12)	52.2 (14.4)	52 (11.2)	51.1 (14.3)	0.911
Females, n (%)	63 (51.2)	60 (48)	207 (50.2)	148 (52.3)	53 (45.3)	208 (53.3)	62 (51.24)	0.873
Sodium (Na)[4], mg	4775.4 (2620.4)	3524.5 (1642.1)	4148 (2018.4)	3372.8 (1472)	4102.2 (1864.3)	4387.7 (2193)	4518.3 (2654.4)	<0.001
Na density (mg/kcal)[4]	2.9 (1.6)	2.3 (1.3)	2.1 (1)	2 (1)	2.3 (1.1)	2.5 (1.5)	2.8 (1.9)	<0.001
Excessive Na[5], n (%)	114 (92.7)	104 (83.2)	383 (93.0)	234 (82.67)	109 (93.2)	345 (88.5)	102 (84.3)	<0.001
Potassium (K)[4], mg	1578.2 (585.6)	1335.6 (554.6)	1625.9 (623.7)	1449.8 (709.5)	1690.3 (758.1)	1687.9 (685)	1379.5 (610.8)	<0.001
K density (mg/kcal)[4]	0.9 (0.4)	0.8 (0.3)	0.8 (0.3)	0.8 (0.3)	0.9 (0.3)	0.9 (0.3)	0.8 (0.3)	<0.001
Deficient K[5], n (%)	122 (99.2)	125 (100)	405 (98.3)	279 (98.6)	113 (96.6)	379 (97.2)	119 (98.35)	0.177
Na/K ratio[4]	3.3 (1.9)	3.1 (1.8)	2.8 (1.4)	2.8 (1.8)	2.7 (1.6)	2.9 (1.9)	3.7 (3)	<0.001

Energy intake[4], kcal	1752.7 (533)	1670.4 (519)	2069.8 (622.6)	1843.4 (533.6)	1894.3 (565.3)	1931.7 (649.2)	1764.9 (558.2)	<0.001
Animal-source foods[4], %kcal	19.1 (11)	22.5 (12.3)	26.1 (12.9)	25.4 (13.2)	14.4 (9.5)	24.7 (11.3)	6.4 (8.1)	<0.001
Fried food intake[4], n (%)	39 (31.7)	18 (14.4)	22 (5.3)	43 (15.2)	36 (30.8)	118 (30.3)	27 (22.3)	<0.001
Ever smoked, n (%)	35 (28.5)	52 (41.6)	174 (42.2)	112 (39.6)	55 (47.0)	164 (42.1)	41 (33.9)	0.117
Alcohol use, n (%)	44 (35.8)	44 (35.2)	126 (30.6)	89 (31.4)	45 (38.5)	98 (25.1)	36 (29.8)	0.002
Pre/probiotics use, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	7 (1.8)	0 (0)	<0.001
NSAIDs use, n (%)	1 (0.8)	0 (0)	1 (0.2)	7 (2.5)	2 (1.7)	7 (1.8)	0 (0)	0.254
PPIs use, n (%)	2 (1.6)	1 (0.8)	2 (0.5)	0 (0)	1 (0.9)	3 (0.8)	0 (0)	0.855
Physical activity[6], n (%)								<0.001
Low	29 (23.6)	49 (39.2)	104 (25.2)	98 (34.6)	39 (33.3)	145 (37.2)	32 (26.4)	
Middle	56 (45.5)	34 (27.2)	132 (32.0)	96 (33.9)	40 (34.2)	122 (31.3)	41 (33.9)	
High	38 (30.9)	42 (33.6)	176 (42.7)	89 (31.5)	38 (32.5)	123 (31.5)	48 (39.8)	
Urbanization[7], n (%)								<0.001
Low	29 (23.6)	51 (40.8)	159 (38.6)	125 (44.2)	17 (14.5)	79 (20.3)	67 (55.4)	
Middle	62 (50.4)	17 (13.6)	75 (18.2)	100 (35.3)	40 (34.2)	202 (51.8)	16 (13.2)	
High	32 (26.0)	57 (45.6)	178 (43.2)	58 (20.5)	60 (51.3)	109 (28.0)	38 (31.4)	

Income[8], n (%)								<0.001
Low	19 (15.5)	56 (44.8)	225 (54.6)	85 (30.0)	27 (23.1)	138 (35.4)	51 (42.2)	
Middle	37 (30.1)	29 (23.2)	133 (32.3)	94 (33.2)	49 (41.9)	136 (34.9)	38 (31.4)	
High	67 (54.5)	40 (32)	54 (13.1)	104 (36.8)	41 (35.0)	116 (29.7)	32 (26.6)	
High school completion, n (%)	52 (42.3)	38 (30.4)	101 (24.5)	80 (28.3)	41 (35.0)	137 (35.1)	38 (31.4)	<0.001
Occupation[9], n (%)								<0.001
Not-working	63 (51.2)	66 (52.8)	162 (39.3)	130 (45.9)	50 (42.7)	222 (56.9)	66 (54.6)	
Agriculture worker	3 (2.4)	12 (9.6)	80 (19.4)	57 (20.1)	18 (15.4)	40 (10.3)	20 (16.5)	
Laborers	28 (22.8)	25 (20)	161 (39.1)	55 (19.4)	34 (29.1)	81 (20.8)	17 (14.1)	
Professional	15 (12.2)	15 (12)	4 (1.0)	20 (7.1)	9 (7.7)	27 (6.9)	10 (8.3)	
Manager	14 (11.4)	7 (5.6)	5 (1.2)	21 (7.4)	6 (5.1)	20 (5.1)	8 (6.6)	
High blood pressure[10], n (%)	68 (55.3)	76 (60.8)	266 (64.6)	166 (58.7)	66 (56.4)	243 (62.3)	64 (52.9)	<0.001
Overweight[10], n (%)	55 (44.7)	53 (42.4)	152 (37.3)	150 (53)	52 (44.8)	189 (48.5)	52 (43.0)	<0.001

Mean (SD), unless noted as n (%); NSAID, non-steroid anti-inflammatory drug; PPI, proton-pump inhibitor.

[1]Provices and megacities were compared using analysis of variance (ANOVA) for continues variables and Chi-squared test for categorical variables.

[2]Shannon index at genus level was calculated using $-\sum p_i ln p_i$, where p_i is the proportional abundance of genera i.

[3]Richness measured the number of distinct genera per subject.

[4]Nutrients intakes estimated by 3-consecutive dietary recalls, household food inventories, and a Chinese food composition table. Sodium density and potassium density were calculated using absolute sodium and potassium intakes divided by total energy intake, respectively.

[5]Excessive sodium (≥2 g) and deficient potassium (<3.5 g) consumption was defined according to the WHO recommendation [115].

[6]Physical activity measured via 7-day recalls in METS/week was categorized to tertiles to represent low (≤40.8 METs/week), medium (40.8-144.5 METs/week), and high (>144.5 METs/week) levels of physical activity.

[7]Urbanization index, a 12-component scale that includes population density, economic activity, transportation infrastructure, sanitation, etc., to define and distinguish urbanicity, was categorized to tertiles to represent low (≤63), medium (63.4-84.3), and high (>84.3) levels of urbanization.

[8]Per capita household income was categorized to year-specific tertiles to represent low (≤9.4k Yuan), medium (9.4-22.4k Yuan), and high (>22.4k Yuan) levels of income.

[9]Occupation was categorized into the following types: unemployed, agriculture worker (e.g. farmer, fisherman, hunter), laborers (e.g. craftsman, logger), professional (e.g. doctor, teacher), manager (e.g. government official, director), and other (e.g. athlete, artist).

[10]High blood pressure was defined as systolic blood pressure or diastolic blood pressure ≥130/80 mmHg, or self-reported high blood pressure. Overweight was defined as BMI (weight/height^2) ≥ 24 kg/m^2. Blood pressure, weight, and height were measured by trained examiners.

Table 4.2. Associations of sodium density, potassium density, and sodium to potassium (Na/K) ratio with within-person gut microbial diversity measures (n=2,833)[1]

	Sodium density		Potassium density		Na/K ratio	
	Coefficient (95% CI)	**P-value**	**Coefficient (95% CI)**	**P-value**	**Coefficient (95% CI)**	**P-value**
Shannon index[2]	0.00 (-0.01, 0.01)	0.554	0.02 (-0.02, 0.05)	0.419	0.00 (-0.00, 0.01)	0.566
Richness[3]	0.38 (-0.68, 1.44)	0.483	-2.28 (-6.72, 2.17)	0.316	0.29 (-0.41, 0.98)	0.418

CI, confidence interval.

[1]Linear regression model was adjusted for age, sex, provinces or megacities, batch or plate run, urbanization, occupation, income, education, total energy intake, %kcal from animal-source foods, fried food intake, physical activity, smoking, alcohol, pro/prebiotic intake in past month, NSAIDs intake in past two weeks, and PPIs intake in past month. Interaction with province/megacity was removed from all models because p-value >0.10.

[2]Shannon index at genus level was calculated using $-\sum p_i ln p_i$, where p_i is the proportional abundance of genera i.

[3]Richness measured the number of distinct genera per subject with rarefication.

Table 4.3. Associations of sodium density, potassium density, and sodium to potassium (Na/K) ratio with between-person gut microbial diversity (n=2,833)[1]

	Sodium density		**Potassium density**		**Na/K ratio**	
	Interaction[2]	**Joint[3]**	**Interaction[2]**	**Joint[3]**	**Interaction[2]**	**Joint[3]**
R-squared[4]	0.07%		0.11%		0.10%	
Sum of squares	0.88	0.93	1.02	1.06	1.18	1.25
F	1.32	1.31	1.53	1.50	1.78	1.76
P value	0.002	0.003	0.001	0.001	0.001	0.001

[1]Model was adjusted for age, sex, provinces or megacities, batch or plate run, urbanization, occupation, income, education, total energy intake, %kcal from animal-source foods, fried food intake, physical activity, smoking, alcohol, pro/prebiotic intake in past month, NSAIDs intake in past two weeks, and PPIs intake in past month.

[2]The interaction of sodium density, potassium density, or Na/K ratio with province/megacity.

[3]The joint tests of sodium density, potassium density, or Na/K ratio with its interaction with province/megacity. Pseudo F statistics and p-values of the interaction and joint test were obtained from partial distance-based redundancy analysis (dbRDA) conditioned on the rest of the model variables followed by an analysis of variance (ANOVA) test with 999 permutations.

[4]R-squared was estimated in a univariate dbRDA model for sodium density, potassium density, or Na/K ratio.

Table 4.4. Associations of sodium density, potassium density, and sodium to potassium (Na/K) ratio with the overall plasma metabolome (n=392)[1]

	Sodium density		**Potassium density**	**Na/K ratio**	
	Interaction[2]	**Joint[3]**		**Interaction[2]**	**Joint[3]**
R-squared[4]	0.53%		0.47%	0.34%	
Sum of squares	0.13	0.24	0.17	0.14	0.24
F	1.55	1.44	2.01	1.70	1.42
P value	0.010	0.009	0.002	0.007	0.007

[1]Model was adjusted for age, sex, provinces or megacities, batch run, urbanization, occupation, income, education, total energy intake, %kcal from animal-source foods, fried food intake, physical activity, smoking, alcohol, pro/prebiotic intake in past month, NSAIDs intake in past two weeks, and PPIs intake in past month. Interaction of potassium density and province/megacity was removed because p-value>0.10.

[2]The interaction between sodium density or Na/K ratio and province/megacity.

[3]The joint test of sodium density or Na/K ratio and its interaction with province/megacity. Pseudo F statistics and p-values of the interaction and joint test were obtained from partial distance-based redundancy analysis (dbRDA) conditioned on the rest of the model variables followed by an analysis of variance (ANOVA) test with 999 permutations.

[4]R-squared was estimated in a univariate dbRDA model for sodium density, potassium density, or Na/K ratio.

Figure 4.1. Heatmap of associations between (A) sodium density, (B) potassium density, and (C) Na/K ratio with specific taxa. N=2,833. Na/K ratio, sodium to potassium ratio; q-value, false discovery rate-adjusted p-value; Interaction, interaction with province/megacity; Joint, joint test of main and interaction effects; "*", unknown genera from family; "**", unknown genera from order; "***" unknown genera from class; "#", q-values<0.10, and "##", q-values<0.05 for province- and megacity-specific estimates; Color and shading of the heatmap indicate the direction and magnitude of model coefficient. Taxa were ordered by joint test q-value and provinces and megacities were ordered by region. Linear regression model was adjusted for age, sex, provinces or megacities, batch or plate run, urbanization, occupation, education, income, total energy intake, %kcal from animal-source foods, fried food intake, physical activity, smoking, alcohol, and intakes of pro/prebiotics, NSAIDs, and PPIs.

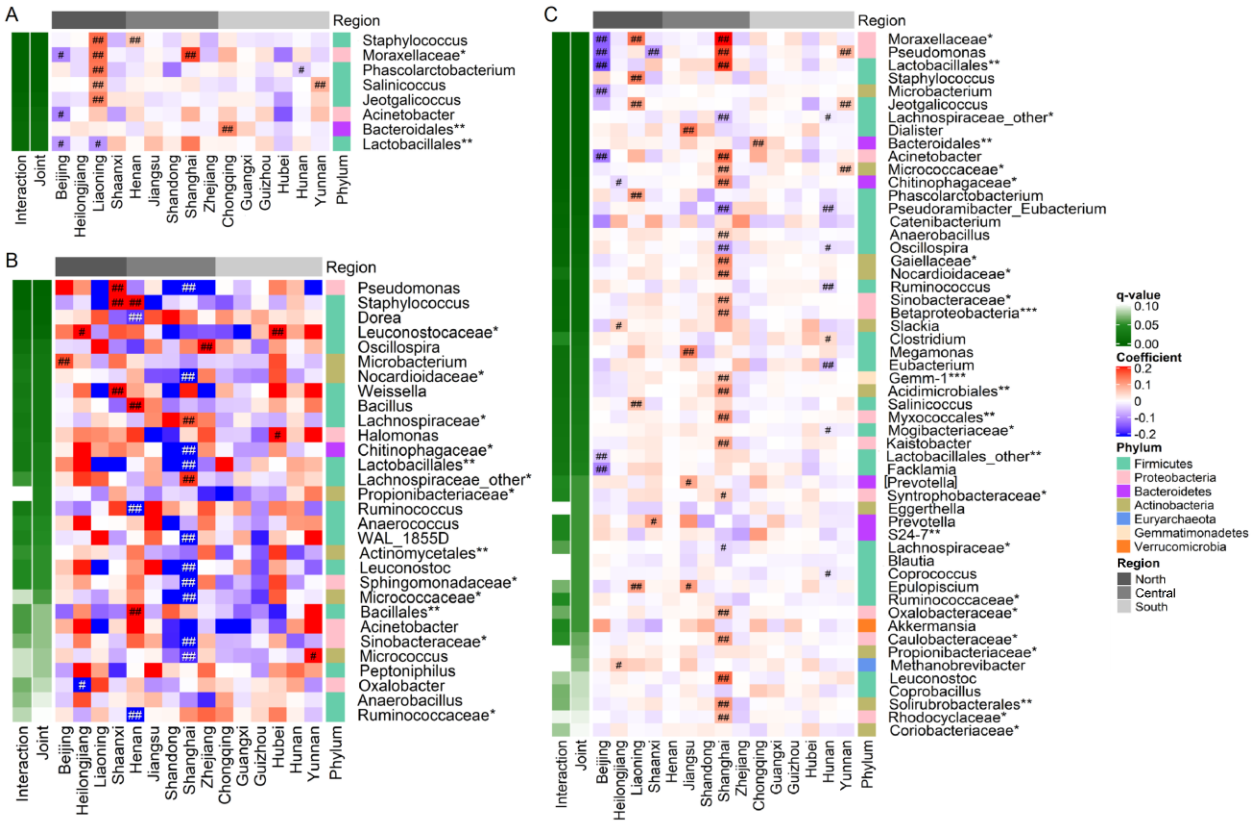


Figure 4.2. Heatmap of associations between (A) sodium density, (B) potassium density, and (C) Na/K ratio with individual metabolites. N=392. Na/K ratio, sodium to potassium ratio; q-value, false discovery rate-adjusted p-value; Interaction, interaction with province/megacity; Joint, joint test of main and interaction effects; "#", q-values<0.10, and "##", q-values<0.05 for province-specific estimates; Color and shading of the heatmap indicate the direction and magnitude of model coefficient. Metabolites were ordered by joint test q-value. Linear regression model was adjusted for age, sex, provinces or megacities, batch or plate run, urbanization, occupation, education, income, total energy intake, %kcal from animal-source foods, fried food intake, physical activity, smoking, alcohol, and intakes of pro/prebiotics, NSAIDs, and PPIs.

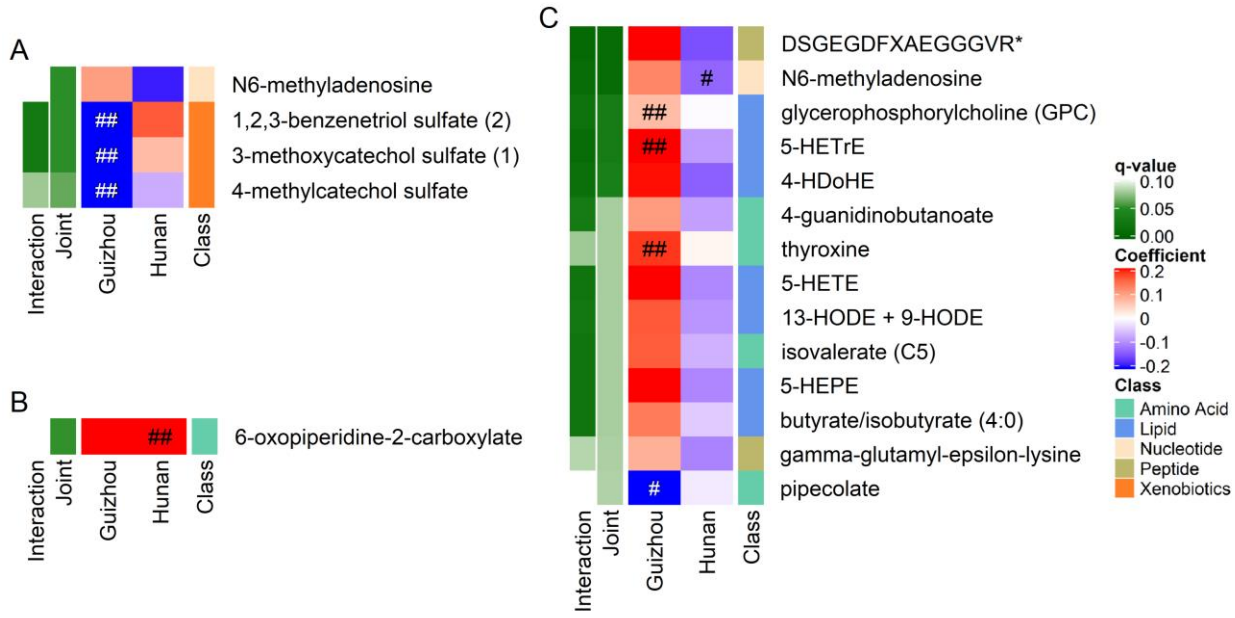


Figure 4.3. Box plots of root mean square errors (RMSEs) of (A) sodium density, (B) potassium density, and (C) Na/K ratio estimated by host factors, microbiota and metabolite data, using random forest regression. N=392. Na/K ratio, sodium to potassium ratio; p-values for 5 iterations of 2-fold cross-validation modified paired t-test that below 0.05 are shown between comparison groups.

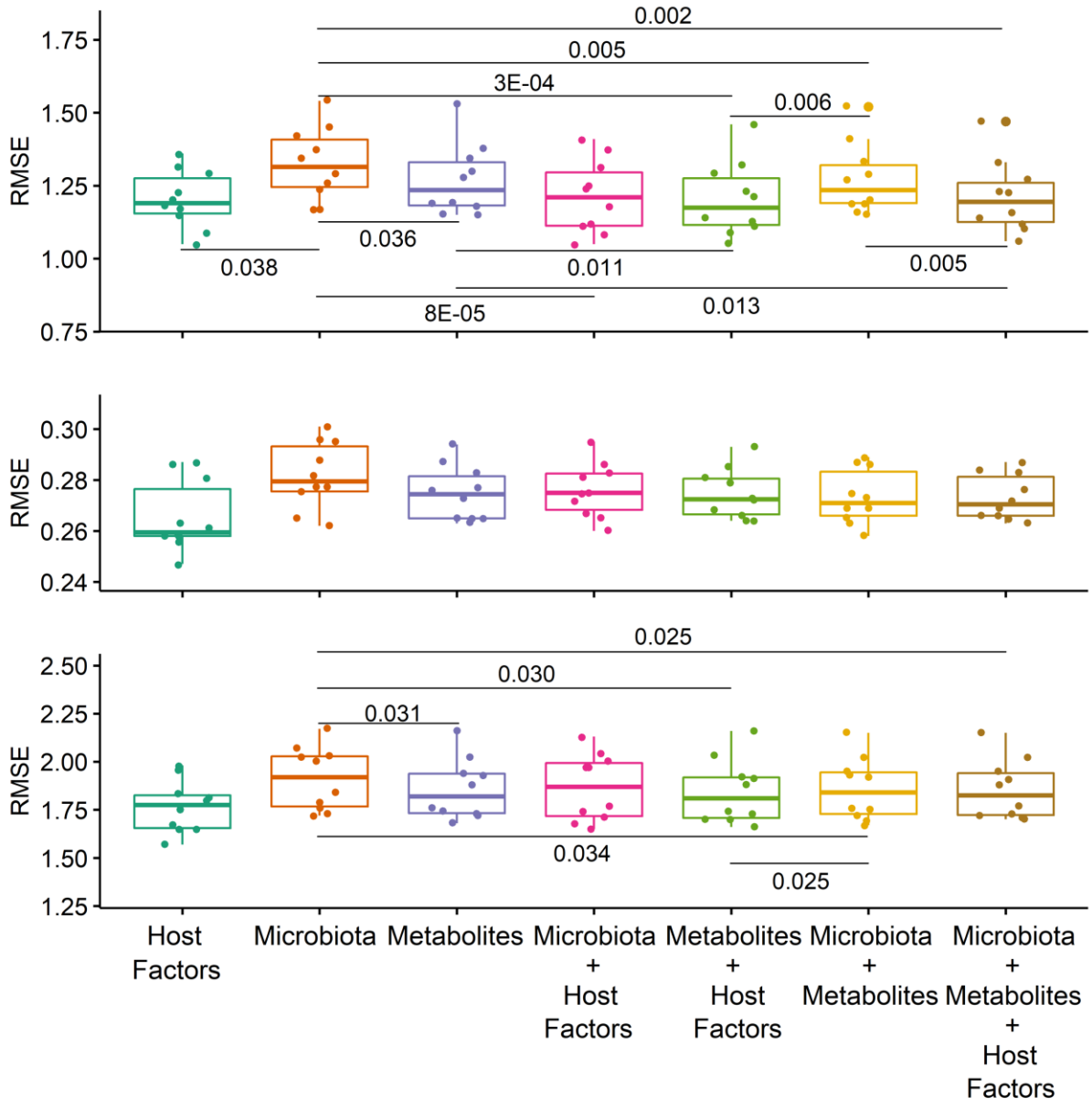


Table S4.1. Characteristics of the plasma metabolome analysis sample

	Total	**Guizhou**	**Hunan**	**P-value[1]**
N	392	133	259	
Shannon index[2]	2.6 (0.3)	2.6 (0.3)	2.6 (0.2)	0.072
Richness[3]	85.8 (20.5)	94 (24.4)	81.5 (16.7)	<0.001
Age, year	51.9 (9)	51.5 (8.9)	52.1 (9.1)	0.488
Females, n (%)	231 (58.9)	82 (61.7)	149 (57.5)	0.498
Sodium (Na)[4], mg	3887.7 (1891.4)	3239 (1373.2)	4220.8 (2031.9)	<0.001
Na density (mg/kcal)[4]	2.2 (1.3)	1.9 (0.9)	2.4 (1.4)	<0.001
Excessive Na[5], n (%)	330 (84.2)	105 (79)	225 (86.9)	0.059
Potassium (K)[4], mg	1562.4 (658)	1370.6 (622.5)	1660.8 (655.1)	0.481
K density (mg/kcal)[4]	0.8 (0.3)	0.8 (0.3)	0.9 (0.3)	<0.001
Deficient K[5], n (%)	385 (98.2)	132 (99.3)	253 (97.7)	0.481
Na/K ratio[4]	2.9 (1.8)	2.9 (1.8)	2.9 (1.9)	0.952
Energy intake[4], kcal	1882.7 (615)	1802.6 (516.8)	1923.8 (656.9)	0.065
Animal-source foods[4], %kcal	25 (12.5)	26.6 (14.2)	24.2 (11.5)	0.071
Fried food intake[4], n (%)	94 (24)	16 (12)	78 (30.1)	<0.001
Ever smoked, n (%)	154 (39.3)	51 (38.4)	103 (39.8)	0.87
Alcohol use, n (%)	100 (25.5)	37 (27.8)	63 (24.3)	0.529
Pre/probiotics use, n (%)	4 (1)	0 (0)	4 (1.5)	0.363
NSAIDs use, n (%)	4 (1)	2 (1.5)	2 (0.8)	0.88
PPIs use, n (%)	2 (0.5)	0 (0)	2 (0.8)	0.789
Physical activity[6], n (%)				0.23
Low	136 (34.69)	40 (30.08)	96 (37.07)	
Middle	127 (32.4)	50 (37.59)	77 (29.73)	
High	129 (32.91)	43 (32.33)	86 (33.2)	
Urbanization[7], n (%)				<0.001
Low	120 (30.6)	66 (49.6)	54 (20.9)	
Middle	178 (45.4)	45 (33.8)	133 (51.4)	

High	94 (24)	22 (16.5)	72 (27.8)	
Income[8], n (%)				0.268
Low	134 (34.2)	40 (30.1)	94 (36.3)	
Middle	136 (34.7)	45 (33.8)	91 (35.1)	
High	122 (31.1)	48 (36.1)	74 (28.6)	
High school completion, n (%)	127 (32.4)	41 (30.8)	86 (33.2)	0.717
Occupation[9], n (%)				0.011
Unemployed	203 (51.8)	58 (43.6)	145 (56)	
Agriculture worker	62 (15.8)	32 (24.1)	30 (11.6)	
Laborers	79 (20.2)	26 (19.6)	53 (20.5)	
Professional	28 (7.1)	12 (9)	16 (6.2)	
Manager	20 (5.1)	5 (3.76)	15 (5.79)	
High blood pressure[10], n (%)	241 (61.5)	78 (58.7)	163 (62.9)	0.474
Overweight[10], n (%)	193 (49.2)	68 (51.1)	125 (48.3)	0.667

Mean (SD), unless noted as n (%); NSAID, non-steroid anti-inflammatory drug; PPI, proton-pump inhibitor.

[1]Provices and megacities were compared using analysis of variance (ANOVA) for continues variables and Chi-squared test for categorical variables.

[2]Shannon index at genus level was calculated using $-\sum p_i ln p_i$, where p_i is the proportional abundance of genera i.

[3]Richness measured the number of distinct genera per subject.

[4]Nutrients intakes estimated by 3-consecutive dietary recalls, household food inventories, and a Chinese food composition table. Sodium density and potassium density were calculated using absolute sodium and potassium intakes divided by total energy intake, respectively.

[5]Excessive sodium (≥2 g) and deficient potassium (<3.5 g) consumption was defined according to the WHO recommendation (35).

[6]Physical activity measured via 7-day recalls in METS/week was categorized to tertiles to represent low (≤40.8 METs/week), medium (40.8-144.5 METs/week), and high (>144.5 METs/week) levels of physical activity.

[7]Urbanization index, a 12-component scale that includes population density, economic activity, transportation infrastructure, sanitation, etc., to define and distinguish urbanicity, was categorized to tertiles to represent low (≤63), medium (63.4-84.3), and high (>84.3) levels of urbanization.

[8]Per capita household income was categorized to year-specific tertiles to represent low (≤9.4k Yuan), medium (9.4-22.4k Yuan), and high (>22.4k Yuan) levels of income.

[9]Occupation was categorized into the following types: unemployed, agriculture worker (e.g. farmer, fisherman, hunter), laborers (e.g. craftsman, logger), professional (e.g. doctor, teacher), manager (e.g. government official, director), and other (e.g. athlete, artist).

[10]High blood pressure was defined as systolic blood pressure or diastolic blood pressure ≥130/80 mmHg, or self-reported high blood pressure. Overweight was defined as BMI (weight/height^2) ≥ 24 kg/m^2. Blood pressure, weight, and height were measured by trained examiners.

Table S4.2. Province/megacity-specific estimates [coefficient (95% CI)] for the associations between sodium density and specific taxa, part 1.

	Beijing	**Heilongjiang**	**Liaoning**	**Shaanxi**	**Henan**	**Jiangsu**	**Shandong**	**Shanghai**
Staphylococcus	-0.01 (-0.05, 0.03)	0.00 (-0.04, 0.04)	0.15 (0.1, 0.19)[2]	-0.06 (-0.12, -0.01)	0.06 (0.03, 0.09)[2]	0.00 (-0.07, 0.06)	-0.04 (-0.1, 0.03)	0.03 (-0.04, 0.09)
Moraxellaceae*	-0.1 (-0.17, -0.04)[1]	-0.03 (-0.09, 0.03)	0.14 (0.07, 0.21)[2]	-0.01 (-0.1, 0.07)	-0.03 (-0.08, 0.02)	0.01 (-0.08, 0.1)	-0.02 (-0.12, 0.09)	0.18 (0.09, 0.27)
Phascolarctobacterium	0.02 (-0.03, 0.07)	-0.03 (-0.08, 0.01)	0.14 (0.08, 0.19)[2]	-0.05 (-0.11, 0.01)	-0.02 (-0.05, 0.02)	0.02 (-0.05, 0.08)	-0.12 (-0.19, -0.04)	-0.01 (-0.08, 0.05)
Salinicoccus	0.00 (-0.03, 0.04)	0.00 (-0.04, 0.03)	0.09 (0.05, 0.14)[2]	-0.03 (-0.08, 0.02)	0.01 (-0.02, 0.04)	0.01 (-0.05, 0.06)	0.01 (-0.05, 0.07)	0.03 (-0.03, 0.08)
Jeotgalicoccus	-0.03 (-0.08, 0.01)	-0.02 (-0.07, 0.02)	0.12 (0.07, 0.17)[2]	0.03 (-0.03, 0.09)	0.03 (-0.01, 0.06)	-0.03 (-0.09, 0.04)	-0.02 (-0.09, 0.05)	-0.01 (-0.07, 0.06)
Acinetobacter	-0.11	0.00	0.06	-0.08	0.04	-0.01	-0.01	0.08

	(-0.17, -0.05)[1]	(-0.06, 0.06)	(-0.01, 0.13)	(-0.17, 0.000)	(-0.01, 0.09)	(-0.1, 0.08)	(-0.12, 0.09)	(-0.01, 0.17)
Bacteroidales**	0.01	0.03	0.02	0.01	-0.03	0.03	0.00	-0.01
	(-0.03, 0.05)	(-0.01, 0.07)	(-0.03, 0.06)	(-0.04, 0.06)	(-0.06, 0.00)	(-0.03, 0.09)	(-0.07, 0.06)	(-0.07, 0.04)
Lactobacillales**	-0.09	-0.01	-0.09	0.06	0.01	0.07	0.02	0.13
	(-0.14, -0.03)[1]	(-0.07, 0.04)	(-0.16, -0.03)[1]	(-0.01, 0.14)	(-0.04, 0.05)	(-0.01, 0.15)	(-0.07, 0.11)	(0.05, 0.22)

Table S4.2. Province/megacity-specific estimates [coefficient (95% CI)] for the associations between sodium density and specific taxa, part 2

	Zhejinag	**Chongqing**	**Guangxi**	**Guizhou**	**Hubei**	**Hunan**	**Yunan**
Staphylococcus	-0.07 (-0.11, -0.02)	0.02 (-0.04, 0.08)	-0.02 (-0.06, 0.02)	-0.03 (-0.07, 0.02)	-0.02 (-0.09, 0.05)	-0.01 (-0.03, 0.02)	-0.02 (-0.06, 0.02)
Moraxellaceae*	-0.05 (-0.12, 0.01)	0.08 (-0.01, 0.16)	0.00 (-0.06, 0.06)	0.03 (-0.04, 0.1)	-0.12 (-0.22, -0.02)	0.03 (-0.01, 0.07)	-0.04 (-0.1, 0.02)
Phascolarctobacterium	0.01 (-0.04, 0.06)	0.02 (-0.04, 0.09)	0.01 (-0.04, 0.05)	-0.03 (-0.08, 0.03)	0.01 (-0.06, 0.09)	-0.05 (-0.08, -0.01)[1]	-0.01 (-0.05, 0.04)
Salinicoccus	-0.02 (-0.06, 0.02)	-0.03 (-0.08, 0.02)	-0.01 (-0.04, 0.03)	-0.05 (-0.09, -0.01)	-0.03 (-0.09, 0.03)	0.00 (-0.03, 0.03)	0.08 (0.05, 0.12)[2]
Jeotgalicoccus	-0.02 (-0.06, 0.03)	0.01 (-0.05, 0.07)	0.01 (-0.03, 0.06)	-0.02 (-0.07, 0.03)	-0.09 (-0.16, -0.02)	0.00 (-0.03, 0.03)	0.04 (0.00, 0.09)
Acinetobacter	-0.01	0.05	-0.01	0.05	-0.15	-0.01	0.06

	(-0.08, 0.06)	(-0.03, 0.14)	(-0.07, 0.05)	(-0.02, 0.12)	(-0.24, -0.05)	(-0.05, 0.03)	(0.00, 0.12)
Bacteroidales**	-0.01	0.12	0.05	-0.02	-0.04	0.00	-0.02
	(-0.06, 0.03)	(0.07, 0.18)[2]	(0.01, 0.09)	(-0.06, 0.03)	(-0.1, 0.02)	(-0.03, 0.02)	(-0.05, 0.02)
Lactobacillales**	0.01	0.01	0.00	0.04	-0.05	0.02	-0.02
	(-0.05, 0.07)	(-0.06, 0.09)	(-0.06, 0.05)	(-0.02, 0.1)	(-0.14, 0.04)	(-0.01, 0.06)	(-0.07, 0.04)

N=2,833. CI, confidence interval; "*", unknown genera from family; "**", unknown genera from order. Only taxa with false discovery rate-adjusted p-value <0.10 for joint test of sodium density and its interaction with province/megacity are shown here. Taxa were ordered by joint test q-values. Taxon relative abundance was $\log_{10}$ transformed. Linear model was adjusted for age, sex, provinces or megacities, batch or plate run, urbanization, occupation, income, high school completion, energy intake, percent energy from animal food, fried food intake, physical activity, ever smoked, alcohol consumption, and use of probiotic, non-steroid anti-inflammatory drug, and proton pump inhibitor. Please see the Supplementary Table 1 in the supplemental files to see results for the full list of taxa.

[1] false discovery rate-adjusted p-value <0.10;

[2] false discovery rate-adjusted p-value <0.05;

Table S4.3. Province/megacity-specific estimates [coefficient (95% CI)] for the associations between potassium density and specific taxa, part 1

	Beijing	**Heilongjiang**	**Liaoning**	**Shaanxi**	**Henan**	**Jiangsu**	**Shandong**	**Shanghai**
Pseudomonas	0.24 (0.05, 0.44)	0.11 (-0.22, 0.44)	-0.28 (-0.66, 0.1)	0.86 (0.5, 1.23)[2]	-0.11 (-0.33, 0.11)	0.02 (-0.29, 0.34)	-0.8 (-1.26, -0.34)	-0.47 (-0.73, -0.21)[2]
Staphylococcus	-0.08 (-0.22, 0.05)	-0.03 (-0.26, 0.21)	-0.26 (-0.52, 0.01)	0.6 (0.34, 0.85)[2]	0.32 (0.16, 0.47)[2]	-0.22 (-0.44, 0)	-0.04 (-0.37, 0.29)	-0.01 (-0.2, 0.17)
Dorea	0.02 (-0.06, 0.1)	0.04 (-0.1, 0.17)	0.16 (0.01, 0.32)	-0.14 (-0.29, 0.01)	-0.17 (-0.26, -0.08)[2]	0.17 (0.04, 0.3)	0.2 (0.01, 0.39)	0.13 (0.02, 0.23)
Leuconostocaceae*	0.13 (-0.07, 0.33)	0.57 (0.23, 0.92)[1]	-0.09 (-0.48, 0.3)	0.16 (-0.21, 0.54)	0.15 (-0.08, 0.38)	-0.02 (-0.35, 0.31)	-0.56 (-1.04, -0.08)	-0.15 (-0.42, 0.12)
Oscillospira	-0.03 (-0.15, 0.08)	-0.04 (-0.23, 0.16)	0.21 (-0.01, 0.43)	-0.11 (-0.32, 0.11)	-0.17 (-0.3, -0.04)	0.12 (-0.06, 0.31)	-0.05 (-0.32, 0.23)	0.12 (-0.03, 0.28)
Microbacterium	0.17	0.06	-0.08	0.14	0.03	-0.04	-0.05	-0.1

	(0.09, 0.25)[2]	(-0.08, 0.2)	(-0.23, 0.08)	(-0.01, 0.3)	(-0.06, 0.12)	(-0.18, 0.09)	(-0.24, 0.15)	(-0.21, 0.01)
Nocardioidaceae*	0.02	0.00	-0.01	-0.03	0.08	-0.15	-0.15	-0.31
	(-0.08, 0.13)	(-0.18, 0.19)	(-0.22, 0.2)	(-0.24, 0.17)	(-0.05, 0.2)	(-0.33, 0.02)	(-0.41, 0.1)	(-0.45, -0.16)[2]
Weissella	-0.02	0.15	-0.33	0.59	0.09	0.04	-0.23	0.19
	(-0.19, 0.16)	(-0.15, 0.45)	(-0.67, 0.01)	(0.26, 0.92)[2]	(-0.11, 0.29)	(-0.25, 0.32)	(-0.65, 0.19)	(-0.04, 0.43)
Bacillus	-0.01	0.05	0.00	-0.13	0.47	0.12	-0.11	-0.06
	(-0.15, 0.14)	(-0.2, 0.3)	(-0.28, 0.29)	(-0.41, 0.14)	(0.31, 0.64)[2]	(-0.12, 0.36)	(-0.47, 0.24)	(-0.26, 0.14)
Lachnospiraceae*	-0.01	-0.05	0.04	0.01	-0.05	0.12	0.23	0.17
	(-0.08, 0.06)	(-0.17, 0.07)	(-0.1, 0.18)	(-0.12, 0.15)	(-0.14, 0.03)	(0.00, 0.23)	(0.06, 0.4)	(0.07, 0.26)[2]
Halomonas	0.05	0.13	0.12	0.09	0.18	-0.32	-0.19	0.04
	(-0.09, 0.19)	(-0.11, 0.36)	(-0.15, 0.39)	(-0.18, 0.35)	(0.03, 0.34)	(-0.54, -0.09)	(-0.52, 0.15)	(-0.15, 0.22)
Chitinophagaceae*	0.02	0.22	0.09	0.13	0.09	0.00	-0.21	-0.31

	(-0.1, 0.13)	(0.03, 0.41)	(-0.12, 0.31)	(-0.08, 0.34)	(-0.04, 0.22)	(-0.18, 0.19)	(-0.48, 0.06)	(-0.46, -0.16)[2]
Lactobacillales**	0.12	0.49	-0.34	-0.25	-0.01	0.09	-0.44	-0.37
	(-0.06, 0.3)	(0.18, 0.8)	(-0.69, 0.01)	(-0.59, 0.09)	(-0.21, 0.2)	(-0.21, 0.38)	(-0.87, -0.01)	(-0.61, -0.12)[2]
Lachnospiraceae_ other*	0.03	0.18	0.16	0.00	-0.07	0.09	-0.07	0.19
	(-0.05, 0.11)	(0.05, 0.32)	(0.01, 0.31)	(-0.14, 0.15)	(-0.15, 0.02)	(-0.04, 0.22)	(-0.26, 0.11)	(0.08, 0.29)[2]
Propionibacteriaceae*	-0.01	-0.01	-0.1	0.09	-0.04	-0.11	-0.03	-0.04
	(-0.1, 0.08)	(-0.16, 0.15)	(-0.28, 0.07)	(-0.08, 0.26)	(-0.14, 0.06)	(-0.26, 0.04)	(-0.25, 0.18)	(-0.16, 0.08)
Ruminococcus	-0.08	-0.12	-0.01	0.18	-0.22	0.21	0.1	0.02
	(-0.19, 0.04)	(-0.32, 0.09)	(-0.24, 0.22)	(-0.04, 0.4)	(-0.35, -0.08)[2]	(0.02, 0.4)	(-0.19, 0.38)	(-0.14, 0.18)
Anaerococcus	0.04	0.2	0.00	0.00	0.05	0.21	-0.2	-0.07
	(-0.05, 0.13)	(0.04, 0.36)	(-0.18, 0.18)	(-0.17, 0.18)	(-0.05, 0.15)	(0.06, 0.36)	(-0.42, 0.02)	(-0.2, 0.05)

WAL_1855D	-0.02 (-0.15, 0.11)	-0.02 (-0.24, 0.21)	0.21 (-0.04, 0.46)	-0.09 (-0.33, 0.16)	0.01 (-0.14, 0.15)	0.15 (-0.06, 0.37)	0.00 (-0.31, 0.31)	-0.33 (-0.51, -0.16)[2]
Actinomycetales**	0.00 (-0.11, 0.11)	0.05 (-0.13, 0.23)	0.03 (-0.18, 0.24)	-0.03 (-0.23, 0.17)	0.13 (0.01, 0.25)	-0.07 (-0.25, 0.1)	-0.09 (-0.35, 0.16)	-0.13 (-0.27, 0.02)
Leuconostoc	0.06 (-0.12, 0.23)	0.19 (-0.1, 0.49)	-0.33 (-0.67, 0.01)	0.00 (-0.33, 0.33)	0.1 (-0.09, 0.3)	0.25 (-0.03, 0.54)	-0.54 (-0.95, -0.12)	-0.41 (-0.64, -0.17)[2]
Sphingomonadaceae*	0.03 (-0.07, 0.14)	0.09 (-0.09, 0.27)	0.00 (-0.21, 0.2)	0.01 (-0.19, 0.21)	0.13 (0.01, 0.25)	-0.02 (-0.19, 0.15)	-0.06 (-0.32, 0.19)	-0.21 (-0.35, -0.06)[2]
Micrococcaceae*	0.03 (-0.09, 0.14)	-0.04 (-0.24, 0.16)	-0.08 (-0.31, 0.14)	0.03 (-0.2, 0.25)	0.11 (-0.02, 0.24)	-0.01 (-0.2, 0.19)	-0.23 (-0.51, 0.05)	-0.31 (-0.47, -0.15)[2]
Bacillales_other**	-0.13 (-0.28, 0.02)	0.12 (-0.14, 0.37)	-0.13 (-0.42, 0.16)	-0.11 (-0.39, 0.17)	0.32 (0.15, 0.49)[2]	-0.07 (-0.32, 0.17)	0.16 (-0.2, 0.51)	0.07 (-0.13, 0.27)

Acinetobacter	0.08 (-0.12, 0.28)	0.2 (-0.15, 0.55)	-0.38 (-0.78, 0.01)	0.04 (-0.35, 0.42)	0.28 (0.05, 0.51)	0.01 (-0.32, 0.35)	-0.18 (-0.67, 0.31)	-0.3 (-0.58, -0.03)
Sinobacteraceae*	-0.03 (-0.14, 0.08)	0.13 (-0.06, 0.32)	0.01 (-0.2, 0.22)	-0.05 (-0.25, 0.16)	0.15 (0.03, 0.28)	0.08 (-0.1, 0.26)	-0.19 (-0.45, 0.08)	-0.25 (-0.4, -0.1)
Micrococcus	0.04 (-0.04, 0.13)	-0.03 (-0.18, 0.12)	0.02 (-0.15, 0.19)	-0.07 (-0.24, 0.09)	0.00 (-0.1, 0.1)	0.01 (-0.13, 0.16)	-0.12 (-0.33, 0.09)	-0.18 (-0.3, -0.07)
Peptoniphilus	-0.08 (-0.19, 0.03)	0.25 (0.05, 0.44)	0.1 (-0.12, 0.32)	-0.18 (-0.4, 0.03)	0.00 (-0.12, 0.13)	0.24 (0.05, 0.42)	0.04 (-0.23, 0.31)	0.01 (-0.14, 0.16)
Oxalobacter	-0.03 (-0.11, 0.05)	-0.24 (-0.37, -0.1)[1]	0.18 (0.02, 0.33)	-0.03 (-0.18, 0.12)	0.06 (-0.03, 0.15)	-0.1 (-0.23, 0.03)	-0.08 (-0.27, 0.11)	0.00 (-0.11, 0.1)
Anaerobacillus	-0.05 (-0.12, 0.03)	0.17 (0.04, 0.3)	-0.03 (-0.18, 0.12)	0.02 (-0.13, 0.16)	0.04 (-0.04, 0.13)	0.07 (-0.05, 0.19)	-0.03 (-0.21, 0.15)	-0.1 (-0.2, 0.00)

Ruminococcaceae*	0.03 (-0.08, 0.13)	-0.07 (-0.25, 0.1)	0.06 (-0.14, 0.26)	-0.02 (-0.21, 0.18)	-0.24 (-0.36, -0.12)[2]	0.01 (-0.16, 0.18)	0.01 (-0.24, 0.26)	0.11 (-0.02, 0.25)

Table S4.3. Province/megacity-specific estimates [coefficient (95% CI)] for the associations between potassium density and specific taxa, part 2

	Zhejinag	**Chongqing**	**Guangxi**	**Guizhou**	**Hubei**	**Hunan**	**Yunan**
Pseudomonas	-0.31 (-0.59, -0.02)	0.00 (-0.3, 0.3)	-0.05 (-0.27, 0.17)	0.00 (-0.22, 0.22)	0.13 (-0.19, 0.45)	0.07 (-0.16, 0.29)	-0.32 (-0.68, 0.04)
Staphylococcus	-0.11 (-0.31, 0.1)	-0.14 (-0.36, 0.07)	-0.07 (-0.22, 0.09)	0.01 (-0.15, 0.17)	-0.07 (-0.3, 0.16)	0.06 (-0.1, 0.22)	-0.02 (-0.28, 0.23)
Dorea	0.01 (-0.11, 0.13)	0.11 (-0.01, 0.23)	0.04 (-0.05, 0.13)	0.09 (0.00, 0.18)	0.03 (-0.11, 0.16)	0.04 (-0.05, 0.14)	0.01 (-0.14, 0.16)
Leuconostocaceae*	-0.14 (-0.44, 0.15)	-0.15 (-0.47, 0.16)	-0.2 (-0.43, 0.02)	0.06 (-0.17, 0.3)	0.58 (0.25, 0.92)[2]	0.09 (-0.14, 0.33)	0.38 (0.00, 0.76)
Oscillospira	0.36 (0.19, 0.53)[2]	0.1 (-0.08, 0.27)	0.04 (-0.09, 0.17)	-0.12 (-0.26, 0.01)	0.1 (-0.09, 0.29)	-0.01 (-0.14, 0.13)	0.07 (-0.14, 0.29)
Microbacterium	-0.04	-0.04	-0.05	-0.05	0.17	-0.02	-0.06

	(-0.16, 0.08)	(-0.16, 0.09)	(-0.14, 0.04)	(-0.14, 0.04)	(0.04, 0.31)	(-0.11, 0.08)	(-0.21, 0.1)
Nocardioidaceae*	0.13	-0.07	-0.1	-0.11	0.17	-0.08	-0.02
	(-0.03, 0.29)	(-0.23, 0.1)	(-0.22, 0.02)	(-0.23, 0.02)	(-0.01, 0.34)	(-0.21, 0.04)	(-0.22, 0.18)
Weissella	-0.32	0.08	-0.03	0.00	0.33	0.06	0.36
	(-0.58, -0.07)	(-0.2, 0.35)	(-0.23, 0.16)	(-0.2, 0.2)	(0.04, 0.62)	(-0.15, 0.26)	(0.03, 0.69)
Bacillus	-0.09	-0.05	-0.04	0.02	0.09	0.01	0.13
	(-0.31, 0.12)	(-0.27, 0.18)	(-0.21, 0.12)	(-0.15, 0.19)	(-0.16, 0.33)	(-0.16, 0.18)	(-0.14, 0.41)
Lachnospiraceae*	0.13	-0.08	0.02	-0.05	-0.07	0.03	-0.07
	(0.02, 0.23)	(-0.19, 0.04)	(-0.07, 0.1)	(-0.14, 0.03)	(-0.19, 0.05)	(-0.05, 0.12)	(-0.2, 0.06)
Halomonas	0.17	-0.04	-0.04	-0.13	0.39	0.05	0.22
	(-0.04, 0.37)	(-0.26, 0.17)	(-0.2, 0.11)	(-0.29, 0.03)	(0.16, 0.62)[1]	(-0.11, 0.21)	(-0.04, 0.49)
Chitinophagaceae*	0.11	-0.01	-0.09	-0.09	0.11	-0.05	0.00

	(-0.06, 0.28)	(-0.18, 0.17)	(-0.22, 0.04)	(-0.22, 0.04)	(-0.08, 0.29)	(-0.18, 0.08)	(-0.22, 0.21)
Lactobacillales**	-0.12	0.24	-0.06	0.05	0.00	0.09	0.07
	(-0.38, 0.15)	(-0.05, 0.52)	(-0.27, 0.14)	(-0.15, 0.26)	(-0.3, 0.3)	(-0.12, 0.3)	(-0.27, 0.41)
Lachnospiraceae_ other*	-0.04	-0.07	0.04	-0.02	-0.03	0.07	0.04
	(-0.16, 0.07)	(-0.19, 0.05)	(-0.05, 0.12)	(-0.11, 0.07)	(-0.16, 0.1)	(-0.02, 0.16)	(-0.11, 0.18)
Propionibacteriaceae*	-0.19	-0.22	-0.12	-0.07	-0.09	0.03	0.13
	(-0.32, -0.05)	(-0.36, -0.08)	(-0.23, -0.02)	(-0.18, 0.03)	(-0.24, 0.06)	(-0.08, 0.13)	(-0.04, 0.3)
Ruminococcus	0.16	0.02	0.1	-0.1	0.13	0.11	-0.13
	(-0.01, 0.34)	(-0.16, 0.21)	(-0.04, 0.23)	(-0.24, 0.03)	(-0.07, 0.32)	(-0.03, 0.25)	(-0.35, 0.1)
Anaerococcus	0.13	-0.01	-0.06	-0.1	-0.03	0.1	0.11
	(0.00, 0.27)	(-0.16, 0.13)	(-0.16, 0.05)	(-0.2, 0.01)	(-0.18, 0.12)	(0.00, 0.21)	(-0.06, 0.28)

WAL_1855D	0.05 (-0.14, 0.25)	-0.02 (-0.22, 0.18)	0.08 (-0.06, 0.23)	-0.19 (-0.34, -0.04)	0.05 (-0.16, 0.27)	0.00 (-0.15, 0.16)	0.29 (0.05, 0.54)
Actinomycetales**	0.15 (-0.01, 0.31)	-0.15 (-0.31, 0.02)	-0.01 (-0.13, 0.11)	-0.15 (-0.27, -0.02)	0.18 (0.01, 0.36)	-0.15 (-0.28, -0.03)	-0.07 (-0.27, 0.13)
Leuconostoc	0.00 (-0.26, 0.26)	-0.01 (-0.28, 0.26)	0.1 (-0.1, 0.3)	-0.16 (-0.36, 0.05)	0.09 (-0.2, 0.38)	-0.02 (-0.23, 0.18)	-0.14 (-0.46, 0.19)
Sphingomonadaceae*	0.13 (-0.02, 0.29)	-0.13 (-0.29, 0.04)	-0.06 (-0.18, 0.06)	-0.14 (-0.26, -0.02)	0.18 (0.00, 0.35)	-0.09 (-0.21, 0.03)	-0.01 (-0.2, 0.19)
Micrococcaceae*	-0.03 (-0.2, 0.15)	-0.06 (-0.25, 0.12)	-0.09 (-0.22, 0.04)	-0.17 (-0.31, -0.03)	0.16 (-0.04, 0.36)	0.00 (-0.14, 0.13)	-0.02 (-0.24, 0.2)
Bacillales_other**	-0.02 (-0.24, 0.2)	-0.05 (-0.28, 0.18)	-0.13 (-0.3, 0.03)	-0.11 (-0.28, 0.07)	-0.19 (-0.44, 0.05)	-0.01 (-0.19, 0.16)	0.26 (-0.02, 0.54)

Acinetobacter	-0.05 (-0.35, 0.25)	-0.33 (-0.65, -0.02)	-0.23 (-0.46, 0)	-0.05 (-0.28, 0.19)	-0.15 (-0.49, 0.19)	0.1 (-0.14, 0.34)	0.4 (0.01, 0.78)
Sinobacteraceae*	0.07 (-0.09, 0.23)	-0.07 (-0.24, 0.11)	-0.06 (-0.19, 0.06)	-0.01 (-0.14, 0.12)	0.08 (-0.1, 0.26)	-0.13 (-0.26, -0.01)	-0.07 (-0.28, 0.13)
Micrococcus	0.04 (-0.09, 0.17)	-0.06 (-0.2, 0.07)	-0.08 (-0.18, 0.02)	-0.08 (-0.19, 0.02)	-0.01 (-0.16, 0.14)	-0.04 (-0.14, 0.07)	0.27 (0.11, 0.44)
Peptoniphilus	0.15 (-0.02, 0.31)	0.00 (-0.18, 0.18)	0.00 (-0.12, 0.13)	-0.09 (-0.22, 0.04)	0.1 (-0.09, 0.29)	0.14 (0.01, 0.28)	0.1 (-0.12, 0.31)
Oxalobacter	0.07 (-0.05, 0.19)	0.04 (-0.09, 0.16)	-0.02 (-0.11, 0.08)	0.02 (-0.07, 0.11)	0.04 (-0.09, 0.17)	0.09 (0, 0.18)	-0.1 (-0.25, 0.05)
Anaerobacillus	-0.18 (-0.3, -0.07)	0.1 (-0.02, 0.21)	-0.03 (-0.11, 0.06)	0.04 (-0.05, 0.13)	0.04 (-0.09, 0.17)	0.01 (-0.08, 0.09)	-0.02 (-0.17, 0.12)

Ruminococcaceae*	0.15 (0, 0.3)	0.12 (-0.04, 0.28)	0.01 (-0.11, 0.13)	0.04 (-0.08, 0.16)	0.13 (-0.04, 0.3)	0.01 (-0.11, 0.14)	-0.05 (-0.25, 0.14)

N=2,833. CI, confidence interval; "*", unknown genera from family; "**", unknown genera from order. Only taxa with false discovery rate-adjusted p-value <0.10 for joint test of potassium density and its interaction with province/megacity are shown here. Taxa were ordered by joint test q-values. Taxon relative abundance was $\log_{10}$ transformed. Linear model was adjusted for age, sex, provinces or megacities, batch or plate run, urbanization, occupation, income, high school completion, energy intake, percent energy from animal food, fried food intake, physical activity, ever smoked, alcohol consumption, and use of probiotic, non-steroid anti-inflammatory drug, and proton pump inhibitor. Please see the Supplementary Table 2 in the supplemental files to see results for the full list of taxa.

[1] false discovery rate-adjusted p-value <0.10;

[2] false discovery rate-adjusted p-value <0.05;

Table S4.4. Province/megacity-specific estimates [coefficient (95% CI)] for the associations between sodium-to-potassium (Na/K) ratio and specific taxa, part 1

	Beijing	Heilongjiang	Liaoning	Shaanxi	Henan	Jiangsu	Shandong	Shanghai
Moraxellaceae*	-0.14 (-0.2, -0.07)²	-0.02 (-0.06, 0.01)	0.14 (0.08, 0.19)²	-0.02 (-0.07, 0.03)	-0.01 (-0.04, 0.01)	-0.01 (-0.07, 0.06)	-0.02 (-0.09, 0.05)	0.27 (0.18, 0.35)²
Pseudomonas	-0.14 (-0.2, -0.07)²	-0.02 (-0.05, 0.02)	0.02 (-0.03, 0.08)	-0.09 (-0.14, -0.04)²	0.02 (-0.01, 0.04)	-0.01 (-0.07, 0.06)	0.00 (-0.07, 0.07)	0.18 (0.09, 0.26)²
Lactobacillales**	-0.15 (-0.21, -0.1)²	-0.01 (-0.04, 0.02)	-0.04 (-0.09, 0.01)	0.06 (0.01, 0.1)	0.00 (-0.02, 0.02)	0.04 (-0.02, 0.1)	0.04 (-0.03, 0.1)	0.18 (0.11, 0.26)²
Staphylococcus	0.05 (0, 0.09)	0.00 (-0.02, 0.03)	0.13 (0.09, 0.16)²	-0.05 (-0.08, -0.02)	0.01 (-0.01, 0.03)	0.02 (-0.02, 0.07)	-0.02 (-0.07, 0.03)	0.03 (-0.03, 0.08)
Microbacterium	-0.09 (-0.12,	-0.02 (-0.03,	0.00 (-0.02,	-0.01 (-0.03,	0.00 (-0.01,	0.00 (-0.03,	0.01 (-0.02,	0.00 (-0.04,

	-0.07)[2]	0)	0.03)	0.01)	0.01)	0.03)	0.04)	0.03)
Jeotgalicoccus	-0.06	-0.01	0.09	0.00	0.01	0.01	0.00	0.00
	(-0.1,	(-0.03,	(0.05,	(-0.03,	(-0.01,	(-0.04,	(-0.05,	(-0.06,
	-0.01)	0.02)	0.14)[2]	0.04)	0.02)	0.05)	0.05)	0.06)
Lachnospiraceae_								
other*	0.00	0.00	-0.01	-0.02	0.02	-0.03	0.03	-0.06
	(-0.02,	(-0.02,	(-0.03,	(-0.04,	(0.01,	(-0.06,	(0.00,	(-0.09,
	0.03)	0.01)	0.01)	-0.01)	0.02)	0.00)	0.05)	-0.03)[2]
Dialister	0.01	0.03	-0.01	0.04	-0.01	0.14	0.06	-0.03
	(-0.03,	(0.00,	(-0.05,	(0.00,	(-0.03,	(0.09,	(0.01,	(-0.09,
	0.06)	0.05)	0.03)	0.07)	0.01)	0.19)[2]	0.11)	0.03)
Bacteroidales**	0.01	0.02	0.01	0.01	-0.01	0.05	0.01	0.02
	(-0.03,	(0.00,	(-0.02,	(-0.02,	(-0.03,	(0.01,	(-0.03,	(-0.04,
	0.05)	0.05)	0.05)	0.04)	0.01)	0.1)	0.06)	0.07)
Acinetobacter	-0.12	0.00	0.08	-0.04	0.01	-0.01	0.01	0.15
	(-0.18,	(-0.04,	(0.02,	(-0.09,	(-0.01,	(-0.07,	(-0.06,	(0.06,
	-0.05)[2]	0.04)	0.13)	0.01)	0.04)	0.06)	0.08)	0.23)[2]
Micrococcaceae*	-0.02	-0.01	0.00	0.00	-0.01	0.00	0.02	0.09

	(-0.06, 0.02)	(-0.03, 0.01)	(-0.03, 0.04)	(-0.03, 0.03)	(-0.02, 0.01)	(-0.04, 0.04)	(-0.02, 0.07)	(0.04, 0.14)[2]
Chitinophagaceae*	0.00	-0.03	-0.01	-0.01	0.00	0.00	0.01	0.12
	(-0.04, 0.03)	(-0.05, -0.01)	(-0.04, 0.02)	(-0.03, 0.02)	(-0.01, 0.02)	(-0.04, 0.04)	(-0.03, 0.05)	(0.07, 0.17)[2]
Phascolarctobacterium	0.03	-0.02	0.1	-0.02	0.00	0.03	-0.07	-0.01
	(-0.02, 0.08)	(-0.05, 0.01)	(0.05, 0.14)[2]	(-0.06, 0.01)	(-0.02, 0.02)	(-0.02, 0.08)	(-0.13, -0.02)	(-0.07, 0.06)
Pseudoramibacter_ Eubacterium	-0.03	0.01	0.02	-0.01	-0.01	-0.07	0.00	-0.1
	(-0.08, 0.02)	(-0.02, 0.03)	(-0.03, 0.06)	(-0.04, 0.03)	(-0.03, 0.01)	(-0.12, -0.02)	(-0.06, 0.05)	(-0.17, -0.04)[2]
Catenibacterium	-0.11	0.07	-0.01	0.09	-0.02	0.05	-0.05	0.03
	(-0.2, -0.01)	(0.02, 0.13)	(-0.09, 0.07)	(0.01, 0.16)	(-0.06, 0.01)	(-0.05, 0.14)	(-0.15, 0.06)	(-0.09, 0.15)
Anaerobacillus	0.01	0.00	0.03	0.03	0.00	0.00	-0.01	0.05
	(-0.01, 0.04)	(-0.02, 0.01)	(0.01, 0.05)	(0.01, 0.05)	(-0.01, 0.01)	(-0.03, 0.02)	(-0.04, 0.02)	(0.02, 0.08)[2]

Oscillospira	0.03 (-0.01, 0.06)	0.00 (-0.02, 0.02)	0.02 (-0.01, 0.05)	0.00 (-0.03, 0.02)	0.00 (-0.01, 0.02)	-0.01 (-0.05, 0.03)	0.03 (-0.02, 0.07)	-0.08 (-0.12, -0.03)[2]
Gaiellaceae*	-0.03 (-0.06, 0.01)	-0.01 (-0.03, 0.01)	0.00 (-0.03, 0.03)	0.02 (0, 0.05)	0.00 (-0.02, 0.01)	0.00 (-0.04, 0.03)	0.02 (-0.02, 0.06)	0.12 (0.07, 0.16)[2]
Nocardioidaceae*	-0.01 (-0.04, 0.03)	0.00 (-0.02, 0.02)	0.03 (0.00, 0.06)	0.02 (0.00, 0.05)	0.00 (-0.01, 0.01)	0.02 (-0.01, 0.06)	0.01 (-0.02, 0.05)	0.11 (0.06, 0.15)[2]
Ruminococcus	0.03 (-0.01, 0.07)	0.01 (-0.01, 0.03)	0.00 (-0.04, 0.03)	-0.03 (-0.06, 0.00)	0.02 (0.00, 0.03)	-0.02 (-0.06, 0.02)	0.00 (-0.04, 0.05)	0.01 (-0.04, 0.06)
Sinobacteraceae*	0.00 (-0.04, 0.03)	-0.02 (-0.04, 0.00)	0.03 (0, 0.06)	0.03 (0.01, 0.06)	-0.01 (-0.02, 0.01)	-0.01 (-0.04, 0.03)	0.03 (-0.01, 0.07)	0.09 (0.04, 0.14)[2]
Betaproteobacteria***	-0.01 (-0.04, 0.03)	-0.01 (-0.03, 0.00)	0.02 (-0.01, 0.04)	0.02 (0.00, 0.04)	0.01 (0.00, 0.02)	-0.01 (-0.04, 0.02)	0.02 (-0.01, 0.06)	0.09 (0.05, 0.14)[2]

Slackia	-0.05 (-0.1, 0.01)	0.05 (0.02, 0.09)	0.01 (-0.05, 0.06)	0.01 (-0.03, 0.05)	0.00 (-0.03, 0.02)	0.03 (-0.03, 0.1)	0.02 (-0.05, 0.09)	0.09 (0.01, 0.17)
Clostridium	0.02 (-0.03, 0.07)	0.01 (-0.02, 0.03)	0.03 (-0.02, 0.07)	0.02 (-0.02, 0.05)	0.01 (-0.01, 0.02)	-0.01 (-0.06, 0.04)	-0.04 (-0.09, 0.02)	-0.01 (-0.07, 0.06)
Megamonas	0.00 (-0.06, 0.06)	-0.01 (-0.05, 0.02)	0.07 (0.01, 0.12)	0.02 (-0.03, 0.06)	-0.01 (-0.04, 0.01)	0.13 (0.07, 0.19)[2]	-0.05 (-0.12, 0.02)	0.03 (-0.05, 0.1)
[Eubacterium]	-0.05 (-0.11, 0.02)	0.02 (-0.01, 0.06)	0.02 (-0.03, 0.07)	0.00 (-0.05, 0.04)	0.01 (-0.02, 0.03)	0.04 (-0.03, 0.1)	-0.01 (-0.08, 0.06)	0.00 (-0.08, 0.08)
Gemm-1***	-0.03 (-0.06, 0.00)	-0.02 (-0.04, 0.00)	0.02 (-0.01, 0.05)	0.03 (0.00, 0.05)	0.00 (-0.01, 0.01)	0.00 (-0.03, 0.03)	0.02 (-0.02, 0.06)	0.07 (0.03, 0.11)[2]
Acidimicrobiales**	-0.03 (-0.07, 0.01)	-0.02 (-0.04, 0.00)	0.02 (-0.01, 0.05)	0.00 (-0.03, 0.03)	0.00 (-0.02, 0.01)	0.01 (-0.03, 0.05)	0.03 (-0.01, 0.07)	0.11 (0.06, 0.16)[2]

Salinicoccus	0.01 (-0.03, 0.05)	-0.01 (-0.03, 0.01)	0.07 (0.03, 0.1)[2]	-0.03 (-0.06, 0.00)	0.00 (-0.02, 0.01)	0.02 (-0.02, 0.06)	0.01 (-0.03, 0.06)	0.04 (-0.01, 0.1)
Myxococcales**	0.00 (-0.03, 0.04)	-0.02 (-0.04, 0.00)	0.02 (-0.01, 0.05)	0.01 (-0.01, 0.04)	0.00 (-0.02, 0.01)	-0.01 (-0.05, 0.02)	0.00 (-0.04, 0.04)	0.11 (0.06, 0.15)[2]
[Mogibacteriaceae]*	0.00 (-0.02, 0.03)	0.01 (-0.01, 0.02)	0.01 (-0.01, 0.03)	0.02 (0.00, 0.04)	0.01 (0.00, 0.02)	0.03 (0.00, 0.06)	0.01 (-0.02, 0.04)	0.01 (-0.02, 0.05)
Kaistobacter	-0.01 (-0.05, 0.02)	-0.01 (-0.03, 0.01)	0.02 (-0.01, 0.05)	0.02 (0.00, 0.05)	0.00 (-0.02, 0.01)	0.01 (-0.03, 0.04)	0.01 (-0.03, 0.05)	0.09 (0.04, 0.14)[2]
Lactobacillales_ other**	-0.05 (-0.08, -0.02)[2]	0.01 (-0.01, 0.02)	0.01 (-0.02, 0.04)	0.02 (0.00, 0.05)	0.00 (-0.01, 0.01)	0.00 (-0.03, 0.04)	0.00 (-0.04, 0.03)	0.04 (0, 0.08)
Facklamia	-0.1 (-0.16,	-0.01 (-0.04,	0.04 (-0.01,	0.05 (0.00,	0.00 (-0.02,	0.01 (-0.05,	-0.01 (-0.07,	0.03 (-0.04,

	-0.05)[2]	0.02)	0.09)	0.09)	0.02)	0.06)	0.05)	0.1)
[Prevotella]	0.02	0.02	-0.02	0.03	0.00	0.09	0.03	-0.01
	(-0.03,	(-0.01,	(-0.06,	(-0.01,	(-0.02,	(0.03,	(-0.04,	(-0.08,
	0.08)	0.06)	0.03)	0.07)	0.02)	0.15)	0.09)	0.06)
Syntrophobacteraceae*	-0.03	-0.01	0.03	0.03	0.00	0.00	0.03	0.06
	(-0.06,	(-0.03,	(0.00,	(0.01,	(-0.02,	(-0.04,	(-0.01,	(0.01,
	0.01)	0.01)	0.06)	0.06)	0.01)	0.03)	0.06)	0.1)[1]
Eggerthella	-0.02	-0.01	-0.04	-0.04	0.01	-0.05	0.00	-0.05
	(-0.06,	(-0.04,	(-0.08,	(-0.07,	(-0.01,	(-0.1,	(-0.04,	(-0.11,
	0.03)	0.01)	0.00)	-0.01)	0.02)	-0.01)	0.05)	0.01)
Prevotella	0.04	0.00	0.04	0.1	0.00	0.11	-0.08	0.00
	(-0.04,	(-0.05,	(-0.03,	(0.04,	(-0.04,	(0.03,	(-0.17,	(-0.11,
	0.12)	0.04)	0.11)	0.16)	0.03)	0.2)	0.02)	0.11)
S24-7*	0.04	0.02	-0.05	0.02	0.00	0.02	0.01	-0.05
	(-0.03,	(-0.02,	(-0.11,	(-0.03,	(-0.03,	(-0.05,	(-0.06,	(-0.14,
	0.11)	0.06)	0.01)	0.07)	0.02)	0.09)	0.09)	0.04)
Lachnospiraceae*	0.00	0.00	-0.01	0.00	0.00	-0.03	-0.01	-0.04
	(-0.02,	(-0.01,	(-0.03,	(-0.02,	(-0.01,	(-0.05,	(-0.04,	(-0.07,

	0.02)	0.02)	0.01)	0.02)	0.01)	-0.01)	0.02)	-0.01)[1]
Blautia	0.01	0.00	-0.02	-0.02	0.00	-0.04	0.01	-0.03
	(-0.01,	(-0.02,	(-0.04,	(-0.04,	(-0.01,	(-0.06,	(-0.02,	(-0.07,
	0.04)	0.01)	0.01)	0.00)	0.01)	-0.01)	0.04)	0.00)
Coprococcus	0.00	-0.01	-0.01	-0.01	0.01	-0.02	0.02	0.00
	(-0.02,	(-0.02,	(-0.03,	(-0.03,	(0.00,	(-0.05,	(-0.01,	(-0.03,
	0.03)	0.01)	0.02)	0.01)	0.02)	0.01)	0.05)	0.04)
Epulopiscium	-0.04	-0.01	0.1	0.01	0.01	0.1	-0.02	-0.01
	(-0.1,	(-0.05,	(0.05,	(-0.04,	(-0.01,	(0.04,	(-0.09,	(-0.09,
	0.03)	0.02)	0.15)	0.06)	0.04)	0.16)[1]	0.05)	0.07)
Ruminococcaceae*	0.02	0.00	-0.01	0.00	0.02	0.00	0.02	-0.03
	(-0.01,	(-0.02,	(-0.04,	(-0.03,	(0,	(-0.03,	(-0.02,	(-0.07,
	0.05)	0.02)	0.02)	0.02)	0.03)	0.04)	0.06)	0.02)
Oxalobacteraceae*	0.00	0.00	0.02	-0.01	0.01	0.01	0.01	0.08
	(-0.03,	(-0.02,	(0.00,	(-0.03,	(0.00,	(-0.02,	(-0.02,	(0.04,
	0.02)	0.01)	0.05)	0.01)	0.02)	0.03)	0.04)	0.11)[2]
Akkermansia	0.1	-0.02	-0.04	-0.07	0.00	0.05	0.08	-0.03
	(0.02,	(-0.06,	(-0.11,	(-0.13,	(-0.03,	(-0.04,	(-0.01,	(-0.13,

	0.18)	0.03)	0.03)	-0.01)	0.03)	0.13)	0.17)	0.07)
Caulobacteraceae*	-0.01	-0.01	0.01	-0.01	0.00	0.00	0.01	0.1
	(-0.04,	(-0.03,	(-0.02,	(-0.03,	(-0.01,	(-0.03,	(-0.02,	(0.05,
	0.03)	0.00)	0.04)	0.02)	0.01)	0.04)	0.05)	0.14)[2]
Propionibacteriaceae*	-0.01	0.02	0.01	0.02	0.00	0.03	-0.02	0.03
	(-0.03,	(0.01,	(-0.02,	(00.00	(-0.01,	(0,	(-0.05,	(-0.01,
	0.02)	0.04)	0.03)	0.04)	0.02)	0.06)	0.01)	0.06)
Methanobrevibacter	0.02	0.06	0.05	0.04	0.00	0.06	-0.01	0.00
	(-0.04,	(0.03,	(0,	(-0.01,	(-0.02,	(-0.01,	(-0.08,	(-0.08,
	0.08)	0.09)[1]	0.1)	0.08)	0.02)	0.12)	0.05)	0.08)
Leuconostoc	-0.01	0.02	-0.01	0.00	-0.02	0.01	0.00	0.14
	(-0.06,	(-0.01,	(-0.06,	(-0.04,	(-0.04,	(-0.05,	(-0.06,	(0.07,
	0.05)	0.05)	0.04)	0.05)	0.00)	0.07)	0.07)	0.22)[2]
Coprobacillus	-0.01	0.01	0.01	-0.03	0.00	0.02	-0.01	-0.04
	(-0.06,	(-0.02,	(-0.04,	(-0.06,	(-0.02,	(-0.03,	(-0.07,	(-0.1,
	0.04)	0.03)	0.05)	0.01)	0.01)	0.07)	0.04)	0.02)
Solirubrobacterales**	-0.02	-0.01	0.02	0.03	-0.01	0.01	0.02	0.11
	(-0.06,	(-0.04,	(-0.02,	(-0.01,	(-0.02,	(-0.04,	(-0.03,	(0.05,

	0.02)	0.01)	0.05)	0.06)	0.01)	0.05)	0.07)	0.16)[2]
Rhodocyclaceae*	0.00	-0.02	-0.01	-0.03	0.00	-0.01	-0.01	0.07
	(-0.04,	(-0.04,	(-0.04,	(-0.05,	(-0.01,	(-0.05,	(-0.05,	(0.03,
	0.03)	0.00)	0.02)	0.00)	0.01)	0.03)	0.03)	0.12)[2]
Coriobacteriaceae*	-0.02	0.02	0.00	-0.01	0.01	0.01	0.02	0.01
	(-0.06,	(0.00,	(-0.04,	(-0.04,	(0.00,	(-0.03,	(-0.03,	(-0.04,
	0.02)	0.05)	0.03)	0.02)	0.03)	0.06)	0.07)	0.07)

Table S4.4. Province/megacity-specific estimates [coefficient (95% CI)] for the associations between sodium-to-potassium (Na/K) ratio and specific taxa, part 2

	Zhejinag	**Chongqing**	**Guangxi**	**Guizhou**	**Hubei**	**Hunan**	**Yunan**
Moraxellaceae*	-0.04 (-0.1, 0.02)	0.04 (-0.02, 0.09)	0.00 (-0.04, 0.04)	0.03 (-0.01, 0.07)	-0.04 (-0.11, 0.03)	0.02 (-0.01, 0.06)	-0.02 (-0.05, 0.02)
Pseudomonas	-0.01 (-0.06, 0.05)	-0.02 (-0.08, 0.04)	0.01 (-0.03, 0.05)	0.01 (-0.02, 0.05)	-0.02 (-0.09, 0.04)	0.00 (-0.03, 0.03)	0.08 (0.04, 0.11)[2]
Lactobacillales**	0.03 (-0.03, 0.08)	-0.03 (-0.08, 0.03)	0.00 (-0.04, 0.03)	0.01 (-0.02, 0.05)	-0.05 (-0.11, 0.02)	0.01 (-0.02, 0.04)	-0.01 (-0.04, 0.02)
Staphylococcus	-0.05 (-0.09, -0.01)	0.04 (0, 0.08)	-0.01 (-0.04, 0.02)	-0.01 (-0.03, 0.02)	-0.01 (-0.06, 0.04)	-0.01 (-0.03, 0.02)	0 (-0.03, 0.02)
Microbacterium	0.00 (-0.03, 0.02)	-0.01 (-0.03, 0.01)	0.01 (0.00, 0.03)	0.01 (-0.01, 0.02)	0.01 (-0.02, 0.04)	0.00 (-0.01, 0.01)	0.01 (-0.01, 0.02)

Jeotgalicoccus	-0.02 (-0.06, 0.02)	0.03 (-0.02, 0.07)	0.01 (-0.02, 0.04)	0.00 (-0.02, 0.03)	-0.07 (-0.12, -0.02)	-0.01 (-0.03, 0.01)	0.06 (0.03, 0.08)[2]
Lachnospiraceae_ other*	-0.02 (-0.04, 0.01)	0.01 (-0.02, 0.03)	0.00 (-0.02, 0.01)	-0.01 (-0.02, 0.01)	0.00 (-0.02, 0.03)	-0.02 (-0.03, -0.01)[1]	-0.01 (-0.03, 0.00)
Dialister	-0.02 (-0.07, 0.02)	-0.01 (-0.06, 0.03)	0.00 (-0.03, 0.03)	0.01 (-0.02, 0.04)	0.05 (0.00, 0.1)	-0.01 (-0.03, 0.01)	-0.01 (-0.03, 0.02)
Bacteroidales**	-0.04 (-0.08, -0.01)	0.09 (0.05, 0.12)[2]	0.04 (0.01, 0.07)	0.00 (-0.03, 0.02)	-0.03 (-0.08, 0.01)	0.00 (-0.02, 0.02)	-0.01 (-0.03, 0.01)
Acinetobacter	-0.01 (-0.06, 0.05)	0.08 (0.02, 0.14)	0.02 (-0.02, 0.06)	0.04 (0.00, 0.08)	-0.06 (-0.13, 0.01)	-0.01 (-0.05, 0.02)	0.02 (-0.02, 0.05)
Micrococcaceae*	-0.01 (-0.04,	-0.01 (-0.05,	0.02 (0,	0.00 (-0.02,	0.02 (-0.02,	0.00 (-0.02,	0.06 (0.04,

	0.02)	0.02)	0.05)	0.02)	0.06)	0.01)	0.08)[2]
Chitinophagaceae*	-0.04	0.01	0.01	0.01	-0.01	0.00	0.00
	(-0.07,	(-0.02,	(-0.01,	(-0.02,	(-0.05,	(-0.01,	(-0.02,
	-0.01)	0.04)	0.03)	0.03)	0.03)	0.02)	0.02)
Phascolarctobacterium	-0.01	0.02	0.00	-0.03	0.00	-0.03	-0.01
	(-0.06,	(-0.03,	(-0.03,	(-0.06,	(-0.05,	(-0.06,	(-0.04,
	0.03)	0.06)	0.03)	0.00)	0.05)	-0.01)	0.02)
Pseudoramibacter_	0.04	0.04	0.00	-0.02	0.00	-0.05	-0.01
Eubacterium							
	(0.00,	(0.00,	(-0.03,	(-0.05,	(-0.05,	(-0.07,	(-0.04,
	0.08)	0.09)	0.04)	0.01)	0.05)	-0.02)[2]	0.01)
Catenibacterium	0.11	-0.05	-0.04	0.03	0.01	-0.05	-0.04
	(0.03,	(-0.13,	(-0.1,	(-0.02,	(-0.09,	(-0.1,	(-0.09,
	0.19)	0.04)	0.02)	0.09)	0.11)	-0.01)	0.02)
Anaerobacillus	0.03	0.00	0.00	-0.01	-0.01	0.00	0.00
	(0.01,	(-0.02,	(-0.01,	(-0.03,	(-0.04,	(-0.01,	(-0.02,
	0.05)	0.03)	0.02)	0.00)	0.01)	0.01)	0.01)
Oscillospira	-0.03	0.03	0.01	0.01	0.00	-0.03	-0.03

	(-0.06, 0.00)	(-0.01, 0.06)	(-0.01, 0.04)	(-0.01, 0.03)	(-0.04, 0.03)	(-0.05, -0.01)²	(-0.05, -0.01)
Gaiellaceae*	-0.01	-0.01	0.01	0.00	0.01	0.01	0.00
	(-0.05, 0.02)	(-0.05, 0.02)	(-0.01, 0.03)	(-0.02, 0.02)	(-0.03, 0.05)	(0.00, 0.03)	(-0.02, 0.02)
Nocardioidaceae*	-0.02	0.00	0.01	0.01	0.00	0.02	0.00
	(-0.05, 0.01)	(-0.03, 0.03)	(-0.02, 0.03)	(-0.02, 0.03)	(-0.04, 0.03)	(0.00, 0.03)	(-0.02, 0.02)
Ruminococcus	-0.04	-0.01	0.01	0.01	0.01	-0.04	0.00
	(-0.07, 0.00)	(-0.04, 0.03)	(-0.01, 0.04)	(-0.01, 0.04)	(-0.03, 0.05)	(-0.05, -0.02)²	(-0.02, 0.02)
Sinobacteraceae*	-0.02	-0.02	0.01	0.01	0.00	0.00	0.00
	(-0.05, 0.02)	(-0.06, 0.01)	(-0.01, 0.03)	(-0.01, 0.03)	(-0.04, 0.04)	(-0.01, 0.02)	(-0.02, 0.02)
Betaproteobacteria***	-0.03	0.00	0.00	0.00	-0.02	0.00	0.00
	(-0.06, 0.00)	(-0.03, 0.03)	(-0.02, 0.02)	(-0.02, 0.02)	(-0.05, 0.02)	(-0.01, 0.02)	(-0.02, 0.02)
Slackia	0.04	-0.03	0.00	0.01	0.07	-0.04	-0.03

	(-0.02, 0.09)	(-0.08, 0.03)	(-0.04, 0.04)	(-0.03, 0.04)	(0.01, 0.13)	(-0.07, -0.01)	(-0.06, 0.01)
Clostridium	0.04	-0.03	0.03	0.01	0.06	0.04	-0.03
	(0.00, 0.08)	(-0.07, 0.01)	(0.00, 0.06)	(-0.02, 0.04)	(0.01, 0.11)	(0.01, 0.06)[1]	(-0.06, 0.00)
Megamonas	-0.01	-0.02	-0.01	-0.01	0.01	0.03	-0.02
	(-0.06, 0.04)	(-0.08, 0.03)	(-0.04, 0.03)	(-0.05, 0.02)	(-0.06, 0.07)	(0, 0.06)	(-0.05, 0.02)
[Eubacterium]	0.07	-0.05	0.02	0.02	0.05	-0.06	0.00
	(0.01, 0.12)	(-0.11, 0)	(-0.02, 0.06)	(-0.02, 0.05)	(-0.01, 0.12)	(-0.09, -0.03)[2]	(-0.04, 0.03)
Gemm-1***	-0.03	-0.01	0.01	0.00	0.00	0.01	0.00
	(-0.06, 0)	(-0.04, 0.02)	(-0.01, 0.03)	(-0.02, 0.02)	(-0.04, 0.03)	(-0.01, 0.02)	(-0.02, 0.02)
Acidimicrobiales**	-0.03	0.00	0.01	0.00	0.00	0.00	0.00
	(-0.07, 0.00)	(-0.04, 0.03)	(-0.02, 0.03)	(-0.03, 0.02)	(-0.04, 0.04)	(-0.02, 0.02)	(-0.02, 0.02)
Salinicoccus	-0.01	-0.01	0.00	-0.01	-0.03	0.00	0.03

	(-0.04, 0.03)	(-0.05, 0.02)	(-0.03, 0.02)	(-0.03, 0.01)	(-0.07, 0.01)	(-0.02, 0.02)	(0.01, 0.05)
Myxococcales**	-0.03	0.00	0.00	-0.01	0.00	0.01	0.01
	(-0.07, 0.00)	(-0.03, 0.03)	(-0.02, 0.03)	(-0.03, 0.02)	(-0.03, 0.04)	(-0.01, 0.02)	(-0.01, 0.03)
[Mogibacteriaceae]*	0.00	0.01	-0.01	-0.01	0.03	-0.02	-0.02
	(-0.02, 0.03)	(-0.02, 0.03)	(-0.03, 0.01)	(-0.03, 0.01)	(0, 0.06)	(-0.03, -0.01)[1]	(-0.03, 0.00)
Kaistobacter	-0.03	-0.01	0.00	0.00	0.02	0.02	0.00
	(-0.06, 0)	(-0.05, 0.02)	(-0.02, 0.02)	(-0.03, 0.02)	(-0.02, 0.06)	(0.00, 0.03)	(-0.02, 0.02)
Lactobacillales_ other**	0.03	-0.01	-0.01	-0.01	-0.04	0.01	0.00
	(0.01, 0.06)	(-0.04, 0.02)	(-0.03, 0.01)	(-0.03, 0.01)	(-0.08, -0.01)	(-0.01, 0.02)	(-0.02, 0.02)
Facklamia	0.01	0.03	0.01	0.02	-0.05	-0.01	0.00
	(-0.04, 0.05)	(-0.02, 0.08)	(-0.03, 0.04)	(-0.01, 0.06)	(-0.11, 0.01)	(-0.04, 0.02)	(-0.03, 0.03)

[Prevotella]	-0.03 (-0.07, 0.02)	0.07 (0.02, 0.12)	0.02 (-0.02, 0.05)	-0.02 (-0.05, 0.02)	0.01 (-0.05, 0.07)	-0.02 (-0.05, 0.01)	-0.01 (-0.04, 0.02)
Syntrophobacteraceae*	-0.03 (-0.06, 0)	-0.01 (-0.04, 0.02)	0.00 (-0.02, 0.03)	0.00 (-0.02, 0.02)	0.03 (0, 0.07)	0.00 (-0.01, 0.02)	0.00 (-0.02, 0.02)
Eggerthella	-0.01 (-0.05, 0.03)	0.03 (-0.01, 0.07)	-0.02 (-0.04, 0.01)	0.01 (-0.02, 0.03)	0.02 (-0.03, 0.06)	-0.02 (-0.05, 0.00)	-0.02 (-0.05, 0.00)
Prevotella	-0.05 (-0.12, 0.02)	-0.02 (-0.09, 0.06)	0.06 (0, 0.11)	-0.02 (-0.07, 0.03)	0.02 (-0.07, 0.1)	-0.02 (-0.06, 0.02)	-0.02 (-0.07, 0.03)
S24-7*	-0.08 (-0.14, -0.02)	0.1 (0.04, 0.17)	0.02 (-0.03, 0.06)	0.00 (-0.04, 0.05)	0.00 (-0.07, 0.08)	-0.04 (-0.08, -0.01)	-0.01 (-0.05, 0.03)
Lachnospiraceae*	-0.03 (-0.05, -0.01)	0.02 (0, 0.04)	0.00 (-0.01, 0.02)	0.00 (-0.01, 0.02)	0.01 (-0.02, 0.03)	-0.01 (-0.02, 0.01)	-0.01 (-0.03, 0.00)

Blautia	0.00 (-0.03, 0.02)	0.00 (-0.03, 0.02)	-0.01 (-0.03, 0.01)	-0.01 (-0.03, 0.01)	-0.02 (-0.05, 0.01)	-0.02 (-0.03, 0.00)	0.00 (-0.02, 0.01)
Coprococcus	-0.02 (-0.04, 0.01)	-0.03 (-0.05, 0.00)	-0.01 (-0.02, 0.01)	0.00 (-0.02, 0.01)	0.02 (-0.01, 0.05)	-0.02 (-0.03, -0.01)[1]	-0.01 (-0.03, 0.00)
Epulopiscium	0.01 (-0.04, 0.07)	-0.01 (-0.07, 0.04)	0.01 (-0.03, 0.05)	0.00 (-0.04, 0.03)	0.04 (-0.02, 0.1)	-0.01 (-0.04, 0.02)	0.01 (-0.02, 0.05)
Ruminococcaceae*	-0.05 (-0.08, -0.02)	0.00 (-0.03, 0.03)	0.00 (-0.02, 0.02)	-0.02 (-0.04, 0.00)	-0.01 (-0.04, 0.03)	-0.01 (-0.03, 0.01)	-0.01 (-0.03, 0.00)
Oxalobacteraceae*	-0.01 (-0.04, 0.01)	0.01 (-0.01, 0.04)	0.01 (-0.01, 0.03)	0.00 (-0.02, 0.01)	0.00 (-0.03, 0.03)	0.00 (-0.01, 0.01)	0.00 (-0.02, 0.01)
Akkermansia	-0.01 (-0.08, 0.06)	0.07 (0, 0.14)	-0.03 (-0.08, 0.02)	0.05 (0, 0.1)	0.00 (-0.08, 0.08)	-0.03 (-0.07, 0.01)	-0.01 (-0.05, 0.03)

Caulobacteraceae*	-0.02 (-0.05, 0.01)	0.01 (-0.02, 0.04)	0.01 (-0.01, 0.03)	-0.01 (-0.03, 0.01)	0.01 (-0.03, 0.04)	-0.01 (-0.03, 0.01)	-0.01 (-0.02, 0.01)
Propionibacteriaceae*	0.03 (0, 0.05)	0.02 (-0.01, 0.04)	0.00 (-0.02, 0.02)	0.01 (0, 0.03)	-0.02 (-0.05, 0.01)	0.00 (-0.01, 0.01)	0.00 (-0.02, 0.01)
Methanobrevibacter	-0.02 (-0.07, 0.04)	0.01 (-0.05, 0.06)	0.02 (-0.01, 0.06)	-0.01 (-0.04, 0.03)	0.04 (-0.02, 0.11)	-0.02 (-0.05, 0.01)	-0.01 (-0.05, 0.02)
Leuconostoc	0.00 (-0.05, 0.05)	-0.01 (-0.06, 0.04)	-0.02 (-0.06, 0.02)	0.01 (-0.03, 0.04)	-0.01 (-0.07, 0.05)	0.02 (-0.01, 0.05)	0.03 (0.00, 0.06)
Coprobacillus	0.00 (-0.04, 0.04)	0.06 (0.01, 0.1)	0.04 (0.01, 0.07)	0.01 (-0.02, 0.03)	0.02 (-0.03, 0.07)	-0.03 (-0.05, -0.01)	-0.01 (-0.04, 0.01)
Solirubrobacterales**	-0.05 (-0.09, -0.01)	0.00 (-0.04, 0.04)	0.01 (-0.02, 0.03)	-0.01 (-0.04, 0.01)	0.01 (-0.03, 0.06)	0.01 (-0.01, 0.03)	0.00 (-0.02, 0.03)

Rhodocyclaceae*	-0.04 (-0.08, -0.01)	0.01 (-0.02, 0.04)	0.01 (-0.02, 0.03)	0.01 (-0.02, 0.03)	-0.02 (-0.06, 0.02)	0.00 (-0.01, 0.02)	0.00 (-0.02, 0.02)
Coriobacteriaceae*	-0.01 (-0.05, 0.02)	-0.04 (-0.07, 0.00)	0.01 (-0.02, 0.04)	-0.01 (-0.04, 0.02)	0.04 (0.00, 0.09)	-0.02 (-0.04, 0.00)	-0.03 (-0.05, 0.00)

N=2,833. CI, confidence interval; "*", unknown genera from family; "**", unknown genera from order; "***", unknown genera from class. Only taxa with false discovery rate-adjusted p-value <0.10 for joint test of Na/K ratio and its interaction with province/megacity are shown here. Taxa were ordered by joint test q-values. Taxon relative abundance was log10 transformed. Linear model was adjusted for age, sex, provinces or megacities, batch or plate run, urbanization, occupation, income, high school completion, energy intake, percent energy from animal food, fried food intake, physical activity, ever smoked, alcohol consumption, and use of probiotic, non-steroid anti-inflammatory drug, and proton pump inhibitor. Please see the Supplementary Table 3 in the supplemental files to see results for the full list of taxa.

1 false discovery rate-adjusted p-value <0.10;

2 false discovery rate-adjusted p-value <0.05;

Table S4.5. Province/megacity-specific estimates [coefficient (95% CI)] for the associations between sodium density and specific operational taxonomic units (OUTs), part 1

	Beijing	**Heilongjiang**	**Liaoning**	**Shaanxi**	**Henan**	**Jiangsu**	**Shandong**	**Shanghai**
Phascolarctobacterium	0.03 (-0.02, 0.08)	-0.04 (-0.08, 0.01)	0.16 (0.1, 0.21)	-0.05 (-0.12, 0.01)	0.01 (-0.03, 0.05)	0.01 (-0.06, 0.08)	-0.11 (-0.19, -0.03)	0.00 (-0.08, 0.07)

Table S4.5. Province/megacity-specific estimates [coefficient (95% CI)] for the associations between sodium density and specific operational taxonomic units (OUTs), part 2

	Zhejinag	**Chongqing**	**Guangxi**	**Guizhou**	**Hubei**	**Hunan**	**Yunan**
Phascolarctobacterium	0.03 (-0.02, 0.09)	0.01 (-0.05, 0.08)	-0.01 (-0.06, 0.04)	-0.01 (-0.06, 0.05)	0.03 (-0.05, 0.10)	-0.03 (-0.06, 0.00)	0.00 (-0.05, 0.04)

N=2,833. CI, confidence interval; Only OTUs with false discovery rate-adjusted p-value <0.10 for joint test of sodium density and its interaction with province/megacity are shown here. OTUs were ordered by joint test q-values. OTU relative abundance was $\log_{10}$ transformed. Linear model was adjusted for age, sex, provinces or megacities, batch or plate run, urbanization, occupation, income, high school completion, energy intake, percent energy from animal food, fried food intake, physical activity, ever smoked, alcohol consumption, and use of probiotic, non-steroid anti-inflammatory drug, and proton pump inhibitor. Please see the Supplementary Table 4 in the supplemental files to see results for the full list of OTUs.

Table S4.6. Province/megacity-specific estimates [coefficient (95% CI)] for the associations between potassium density and specific operational taxonomic units (OUTs), part 1

	Beijing	**Heilongjiang**	**Liaoning**	**Shaanxi**	**Henan**	**Jiangsu**	**Shandong**	**Shanghai**
Dorea	0.04 (-0.04, 0.12)	0.07 (-0.08, 0.21)	0.14 (-0.02, 0.3)	-0.17 (-0.33, -0.01)	-0.14 (-0.23, -0.04)	0.18 (0.04, 0.32)	0.24 (0.04, 0.44)	0.16 (0.05, 0.27)
Ruminococcaceae-UCG-014	-0.15 (-0.36, 0.06)	-0.31 (-0.67, 0.05)	-0.34 (-0.75, 0.07)	-0.08 (-0.48, 0.32)	-0.09 (-0.33, 0.15)	0.84 (0.5, 1.19)[2]	0.17 (-0.33, 0.68)	-0.09 (-0.38, 0.19)
Weissella	0.25 (0.03, 0.46)	0.49 (0.12, 0.86)	-0.09 (-0.51, 0.34)	0.17 (-0.24, 0.58)	0.11 (-0.14, 0.36)	-0.03 (-0.39, 0.33)	-0.6 (-1.13, -0.08)	-0.32 (-0.61, -0.03)
Bacteroides	0.07 (-0.21, 0.35)	-0.41 (-0.89, 0.06)	0.26 (-0.28, 0.8)	0.00 (-0.52, 0.53)	0.27 (-0.04, 0.59)	-0.79 (-1.24, -0.33)[2]	0.25 (-0.42, 0.91)	-0.32 (-0.69, 0.05)
Lactobacillus	-0.08 (-0.3, 0.14)	0.27 (-0.1, 0.65)	-0.11 (-0.54, 0.32)	-0.9 (-1.31, -0.48)	0.17 (-0.08, 0.42)	0.00 (-0.36, 0.36)	-0.34 (-0.87, 0.19)	-0.37 (-0.66, -0.07)

Fusicatenibacter	-0.06 (-0.16, 0.05)	-0.16 (-0.34, 0.02)	0.06 (-0.15, 0.26)	0.15 (-0.05, 0.34)	-0.04 (-0.16, 0.08)	0.11 (-0.07, 0.28)	0.40 (0.15, 0.65)	0.15 (0.01, 0.29)

Table S4.6. Province/megacity-specific estimates [coefficient (95% CI)] for the associations between potassium density and specific operational taxonomic units (OUTs), part 2

	Zhejinag	**Chongqing**	**Guangxi**	**Guizhou**	**Hubei**	**Hunan**	**Yunan**
Dorea	0.07 (-0.06, 0.19)	0.14 (0.01, 0.27)	0.07 (-0.02, 0.17)	0.04 (-0.06, 0.14)	0.06 (-0.08, 0.19)	0.06 (-0.04, 0.15)	-0.07 (-0.23, 0.08)
Ruminococcaceae-UCG-014	-0.12 (-0.43, 0.19)	0.2 (-0.13, 0.53)	0.06 (-0.18, 0.3)	-0.1 (-0.34, 0.15)	0.52 (0.17, 0.87)	-0.04 (-0.29, 0.21)	0.19 (-0.21, 0.59)
Weissella	-0.14 (-0.46, 0.18)	-0.24 (-0.58, 0.1)	-0.17 (-0.42, 0.07)	0.11 (-0.15, 0.36)	0.64 (0.28, 1)[2]	0.13 (-0.12, 0.39)	0.33 (-0.08, 0.74)
Bacteroides	0.14 (-0.27, 0.55)	-0.06 (-0.49, 0.37)	0.48 (0.16, 0.79)	-0.13 (-0.45, 0.19)	0.51 (0.05, 0.97)	-0.16 (-0.49, 0.16)	0.52 (-0.01, 1.04)
Lactobacillus	-0.08 (-0.41, 0.24)	-0.03 (-0.37, 0.31)	0.07 (-0.18, 0.32)	-0.19 (-0.44, 0.07)	-0.37 (-0.73, 0.00)	0.27 (0.02, 0.53)	0.14 (-0.28, 0.55)

Fusicatenibacter	0.23 (0.07, 0.38)	-0.15 (-0.31, 0.02)	-0.02 (-0.14, 0.09)	-0.03 (-0.15, 0.09)	-0.09 (-0.26, 0.08)	0.04 (-0.08, 0.16)	-0.05 (-0.24, 0.15)

N=2,833. CI, confidence interval; Only OTUs with false discovery rate-adjusted p-value <0.10 for joint test of potassium density and its interaction with province/megacity are shown here. OTUs were ordered by joint test q-values. OTU relative abundance was $\log_{10}$ transformed. Linear model was adjusted for age, sex, provinces or megacities, batch or plate run, urbanization, occupation, income, high school completion, energy intake, percent energy from animal food, fried food intake, physical activity, ever smoked, alcohol consumption, and use of probiotic, non-steroid anti-inflammatory drug, and proton pump inhibitor. Please see the Supplementary Table 5 in the supplemental files to see results for the full list of OTUs.

[1] false discovery rate-adjusted p-value <0.10;

[2] false discovery rate-adjusted p-value <0.05;

Table S4.7. Province/megacity-specific estimates [coefficient (95% CI)] for the associations between sodium-to-potassium (Na/K) ratio and specific operational taxonomic units (OUTs), part 1

	Beijing	**Heilongjiang**	**Liaoning**	**Shaanxi**	**Henan**	**Jiangsu**	**Shandong**	**Shanghai**
Faecalitalea	0.07 (0.01, 0.13)	0.06 (0.02, 0.09)[1]	-0.01 (-0.06, 0.05)	0.02 (-0.02, 0.06)	0.01 (-0.01, 0.03)	0.08 (0.02, 0.14)	-0.07 (-0.13, 0.00)	-0.08 (-0.15, 0.00)
Ruminococcus_2	0.20 (0.1, 0.31)[2]	0.02 (-0.03, 0.08)	-0.03 (-0.12, 0.06)	0.02 (-0.06, 0.09)	-0.02 (-0.06, 0.02)	0.00 (-0.11, 0.1)	0.03 (-0.08, 0.14)	0.17 (0.04, 0.3)
CAG-56	0.11 (0.03, 0.19)	0.05 (0.01, 0.10)	0.00 (-0.07, 0.07)	-0.08 (-0.14, -0.02)	-0.02 (-0.05, 0.01)	-0.08 (-0.16, 0.00)	0.02 (-0.07, 0.11)	-0.04 (-0.15, 0.06)
Lachnospiraceae_ FCS020_group	0.15 (0.08, 0.22)[2]	0.02 (-0.02, 0.06)	-0.02 (-0.08, 0.04)	-0.01 (-0.07, 0.04)	0.01 (-0.02, 0.04)	0.04 (-0.03, 0.11)	-0.06 (-0.14, 0.02)	0.00 (-0.09, 0.09)
Blautia	0.03 (-0.03, 0.1)	-0.02 (-0.05, 0.02)	-0.01 (-0.07, 0.05)	0.01 (-0.04, 0.05)	0.02 (-0.01, 0.04)	-0.03 (-0.1, 0.03)	0.05 (-0.03, 0.12)	-0.14 (-0.23, -0.06)[2]

Lactobacillus	0.05 (-0.04, 0.13)	0.09 (0.04, 0.13)[2]	-0.04 (-0.11, 0.04)	0.01 (-0.06, 0.07)	0.00 (-0.03, 0.03)	0.06 (-0.03, 0.15)	0.05 (-0.05, 0.14)	0.00 (-0.11, 0.11)
Holdemanella	-0.11 (-0.21, 0.00)	0.04 (-0.02, 0.1)	0.03 (-0.06, 0.12)	0.00 (-0.08, 0.08)	-0.01 (-0.05, 0.03)	0.08 (-0.03, 0.19)	0.02 (-0.09, 0.14)	-0.23 (-0.37, -0.09)
Phascolarctobacterium	0.02 (-0.03, 0.07)	-0.03 (-0.06, 0.00)	0.10 (0.05, 0.14)[2]	-0.03 (-0.07, 0.01)	0.00 (-0.02, 0.02)	0.01 (-0.05, 0.06)	-0.07 (-0.13, -0.01)	0.00 (-0.07, 0.07)
Megamonas	0.01 (-0.06, 0.08)	-0.01 (-0.05, 0.02)	0.08 (0.02, 0.14)	0.03 (-0.02, 0.08)	-0.02 (-0.04, 0.01)	0.16 (0.08, 0.23)[2]	-0.06 (-0.13, 0.02)	0.00 (-0.09, 0.09)
Ruminococcaceae UCG-014	0.08 (0.00, 0.15)	0.04 (0.00, 0.08)	-0.1 (-0.17, -0.03)	0.01 (-0.05, 0.07)	0.01 (-0.02, 0.04)	-0.05 (-0.13, 0.03)	0.03 (-0.06, 0.12)	0.01 (-0.09, 0.11)
Bifidobacterium	0.00 (-0.11, 0.11)	0.08 (0.02, 0.14)	-0.11 (-0.21, -0.01)	0.08 (-0.01, 0.16)	0.00 (-0.04, 0.05)	0.2 (0.08, 0.31)[2]	0.02 (-0.11, 0.14)	-0.08 (-0.23, 0.06)

Catenibacterium	-0.1 (-0.2, -0.01)	0.07 (0.02, 0.13)	-0.02 (-0.11, 0.07)	0.07 (0.00, 0.15)	-0.02 (-0.06, 0.01)	0.05 (-0.05, 0.15)	-0.05 (-0.16, 0.06)	0.05 (-0.08, 0.18)
Holdemanella	0.13 (0.03, 0.23)	0.04 (-0.02, 0.09)	0.01 (-0.07, 0.1)	0.02 (-0.05, 0.1)	0.00 (-0.03, 0.04)	0.04 (-0.06, 0.14)	-0.03 (-0.14, 0.09)	0.18 (0.05, 0.31)
Lachnospira	0.01 (-0.05, 0.07)	-0.04 (-0.07, -0.01)	0.1 (0.05, 0.15)[2]	-0.02 (-0.06, 0.03)	0.00 (-0.03, 0.02)	0.07 (0.01, 0.13)	0.05 (-0.02, 0.11)	0.01 (-0.07, 0.09)
Ruminococcaceae UCG-014	-0.01 (-0.08, 0.07)	0.01 (-0.03, 0.05)	0.13 (0.06, 0.19)[2]	0.02 (-0.04, 0.07)	-0.01 (-0.03, 0.02)	-0.1 (-0.18, -0.03)	-0.09 (-0.17, -0.01)	-0.01 (-0.1, 0.09)
Coprococcus_2	0.06 (-0.01, 0.13)	0.00 (-0.04, 0.04)	-0.01 (-0.07, 0.05)	-0.05 (-0.1, 0.00)	-0.03 (-0.05, 0.00)	0.1 (0.03, 0.17)	0.02 (-0.05, 0.1)	0.17 (0.07, 0.26)[2]
Clostridium_sensu_stricto_1	0.04 (-0.04, 0.12)	0.00 (-0.05, 0.04)	0.13 (0.06, 0.2)[1]	0.05 (-0.01, 0.11)	0.00 (-0.03, 0.03)	0.11 (0.03, 0.19)	-0.04 (-0.13, 0.05)	0.05 (-0.06, 0.15)

Lachnospiraceae*	-0.03 (-0.11, 0.05)	0.00 (-0.05, 0.04)	-0.01 (-0.08, 0.06)	0.00 (-0.06, 0.06)	0.01 (-0.02, 0.04)	0.00 (-0.08, 0.09)	0.12 (0.03, 0.2)	0.06 (-0.05, 0.16)
Faecalibacterium	0.06 (0.00, 0.13)	0.02 (-0.02, 0.05)	0.01 (-0.05, 0.06)	-0.01 (-0.06, 0.03)	-0.02 (-0.04, 0.00)	0.01 (-0.06, 0.07)	0.00 (-0.07, 0.07)	0.02 (-0.06, 0.1)
Lactobacillus	0.05 (-0.05, 0.14)	0.11 (0.06, 0.17)[2]	-0.02 (-0.1, 0.07)	0.02 (-0.05, 0.1)	0.00 (-0.04, 0.04)	0.11 (0.01, 0.21)	0.06 (-0.05, 0.17)	-0.05 (-0.17, 0.08)
Subdoligranulum	0.06 (0.01, 0.11)	0.00 (-0.03, 0.02)	-0.04 (-0.08, 0.00)	-0.01 (-0.05, 0.02)	0.01 (0.00, 0.03)	-0.01 (-0.06, 0.04)	0.00 (-0.06, 0.05)	-0.05 (-0.12, 0.01)
Holdemanella	-0.13 (-0.23, -0.04)	0.06 (0.01, 0.11)	0.01 (-0.07, 0.1)	-0.02 (-0.09, 0.06)	0.01 (-0.03, 0.04)	0.05 (-0.05, 0.15)	-0.03 (-0.14, 0.07)	-0.02 (-0.14, 0.11)
Lachnospiraceae_ ND3007_group	-0.02 (-0.07, 0.03)	0.01 (-0.02, 0.04)	0.03 (-0.01, 0.08)	-0.01 (-0.05, 0.03)	-0.01 (-0.03, 0.01)	0.04 (-0.02, 0.09)	-0.02 (-0.08, 0.04)	0.05 (-0.02, 0.12)

Fusicatenibacter	0.01 (-0.03, 0.04)	0.01 (-0.01, 0.03)	-0.01 (-0.04, 0.02)	0.00 (-0.02, 0.03)	0.00 (-0.01, 0.01)	-0.04 (-0.07, 0.00)	-0.03 (-0.06, 0.01)	-0.07 (-0.12, -0.03)[1]
Subdoligranulum	0.03 (-0.01, 0.07)	0.02 (-0.01, 0.04)	-0.01 (-0.05, 0.03)	0.01 (-0.02, 0.04)	0.01 (-0.01, 0.02)	0.01 (-0.03, 0.06)	0.00 (-0.05, 0.05)	-0.03 (-0.08, 0.03)
Akkermansia	0.12 (0.04, 0.21)	0.00 (-0.05, 0.05)	-0.03 (-0.1, 0.05)	-0.03 (-0.1, 0.03)	0.00 (-0.03, 0.03)	0.06 (-0.03, 0.15)	0.08 (-0.01, 0.18)	-0.06 (-0.17, 0.05)
Ruminococcaceae_ UCG-014	0.13 (0.05, 0.21)	0.02 (-0.03, 0.06)	-0.02 (-0.09, 0.05)	-0.05 (-0.1, 0.01)	0.00 (-0.03, 0.03)	0.05 (-0.03, 0.13)	0.01 (-0.07, 0.09)	0.00 (-0.09, 0.1)
Ruminococcaceae_ UCG-004	0.07 (0.01, 0.13)	-0.01 (-0.05, 0.02)	0.03 (-0.02, 0.09)	-0.02 (-0.06, 0.03)	0.01 (-0.01, 0.03)	-0.01 (-0.07, 0.05)	0.02 (-0.04, 0.09)	-0.14 (-0.22, -0.06)[1]
Eubacterium	0.00 (-0.05, 0.06)	0.00 (-0.03, 0.04)	-0.01 (-0.06, 0.04)	-0.01 (-0.05, 0.03)	0.00 (-0.02, 0.02)	-0.03 (-0.08, 0.03)	0.00 (-0.06, 0.07)	-0.1 (-0.17, -0.02)

Family_XIII*	-0.01	-0.02	-0.01	0.04	0.00	-0.02	0.02	0.11
	(-0.06,	(-0.04,	(-0.05,	(0.01,	(-0.02,	(-0.07,	(-0.04,	(0.05,
	0.03)	0.01)	0.03)	0.08)	0.02)	0.03)	0.07)	0.17)
Clostridium_sensu_	0.00	0.00	0.02	-0.01	-0.01	0.02	-0.05	-0.01
stricto_1	(-0.08,	(-0.05,	(-0.06,	(-0.08,	(-0.04,	(-0.07,	(-0.14,	(-0.12,
	0.09)	0.04)	0.09)	0.05)	0.02)	0.1)	0.04)	0.1)
Christensenellaceae_	0.1	0.01	-0.1	0.01	-0.01	-0.05	0.09	-0.02
R-7_group	(0.03,	(-0.03,	(-0.16,	(-0.04,	(-0.04,	(-0.12,	(0.01,	(-0.12,
	0.17)	0.05)	-0.04)	0.06)	0.02)	0.02)	0.17)	0.07)
Granulicatella	0.1	-0.01	0.00	0.02	0.00	-0.01	-0.02	0.07
	(0.05,	(-0.04,	(-0.05,	(-0.02,	(-0.02,	(-0.06,	(-0.08,	(0,
	0.15)	0.02)	0.04)	0.06)	0.02)	0.05)	0.04)	0.13)
Lachnospiraceae*	0.05	0.09	0.00	0.01	0.00	0.01	-0.04	0.09
	(-0.02,	(0.05,	(-0.07,	(-0.04,	(-0.03,	(-0.06,	(-0.12,	(-0.01,
	0.12)	0.13)[2]	0.06)	0.07)	0.03)	0.09)	0.04)	0.18)
Ruminococcus_1	0.13	-0.01	-0.03	-0.01	0.00	-0.02	-0.01	0.09
	(0.07,	(-0.05,	(-0.08,	(-0.06,	(-0.03,	(-0.09,	(-0.08,	(0,
	0.2)[2]	0.03)	0.03)	0.04)	0.02)	0.05)	0.07)	0.17)

Mogibacterium	-0.07 (-0.13, 0.00)	-0.02 (-0.05, 0.02)	0.07 (0.02, 0.13)	0.03 (-0.02, 0.08)	0.01 (-0.01, 0.04)	0.1 (0.03, 0.17)[1]	0.00 (-0.08, 0.07)	-0.06 (-0.14, 0.03)

Table S4.7. Province/megacity-specific estimates [coefficient (95% CI)] for the associations between sodium-to-potassium (Na/K) ratio and specific operational taxonomic units (OUTs), part 2

	Zhejinag	**Chongqing**	**Guangxi**	**Guizhou**	**Hubei**	**Hunan**	**Yunan**
Faecalitalea	0.00 (-0.05, 0.05)	0.00 (-0.06, 0.05)	0.07 (0.04, 0.11)	0.03 (0, 0.07)	-0.02 (-0.08, 0.05)	0.00 (-0.03, 0.03)	-0.01 (-0.04, 0.02)
Ruminococcus_2	0.02 (-0.07, 0.11)	-0.01 (-0.11, 0.08)	-0.01 (-0.07, 0.06)	0.01 (-0.05, 0.08)	-0.04 (-0.15, 0.07)	-0.08 (-0.13, -0.04)[2]	0.04 (-0.02, 0.09)
CAG-56	-0.11 (-0.18, -0.03)	-0.03 (-0.11, 0.04)	-0.02 (-0.07, 0.03)	0.02 (-0.03, 0.07)	-0.03 (-0.12, 0.05)	-0.02 (-0.06, 0.02)	0.01 (-0.04, 0.05)
Lachnospiraceae_ FCS020_group	-0.09 (-0.15, -0.03)	-0.01 (-0.07, 0.06)	0.01 (-0.03, 0.06)	-0.03 (-0.08, 0.01)	0.05 (-0.02, 0.13)	-0.02 (-0.06, 0.01)	-0.01 (-0.05, 0.03)
Blautia	0.01 (-0.05, 0.07)	-0.01 (-0.07, 0.05)	0.02 (-0.02, 0.07)	-0.07 (-0.11, -0.02)	0.01 (-0.06, 0.08)	-0.05 (-0.08, -0.02)[2]	0.00 (-0.04, 0.03)

Lactobacillus	-0.03 (-0.1, 0.05)	0.12 (0.05, 0.2)	0.00 (-0.05, 0.06)	-0.02 (-0.08, 0.03)	0.06 (-0.03, 0.15)	-0.02 (-0.06, 0.02)	-0.06 (-0.11, -0.01)
Holdemanella	0.08 (-0.01, 0.18)	0.03 (-0.07, 0.13)	0.08 (0.01, 0.15)	0.06 (0, 0.13)	0.04 (-0.07, 0.15)	-0.06 (-0.11, -0.01)	0.01 (-0.05, 0.06)
Phascolarctobacterium	0.01 (-0.04, 0.05)	0.01 (-0.04, 0.05)	-0.01 (-0.04, 0.02)	-0.03 (-0.06, 0.01)	0.01 (-0.04, 0.07)	-0.02 (-0.04, 0.01)	0.00 (-0.03, 0.02)
Megamonas	0.00 (-0.06, 0.06)	-0.01 (-0.08, 0.05)	-0.01 (-0.05, 0.04)	0.00 (-0.05, 0.04)	-0.02 (-0.09, 0.06)	0.03 (0.00, 0.06)	-0.01 (-0.05, 0.03)
Ruminococcaceae UCG-014	0.07 (0.00, 0.14)	-0.05 (-0.12, 0.03)	0.05 (0.00, 0.1)	0.06 (0.02, 0.11)	-0.01 (-0.09, 0.08)	-0.01 (-0.05, 0.02)	0.04 (0.00, 0.08)
Bifidobacterium	-0.07 (-0.17, 0.03)	-0.07 (-0.18, 0.03)	0.03 (-0.04, 0.1)	-0.06 (-0.12, 0.01)	-0.01 (-0.13, 0.11)	0.01 (-0.05, 0.06)	-0.04 (-0.11, 0.02)

Catenibacterium	0.11 (0.02, 0.19)	-0.03 (-0.12, 0.06)	-0.04 (-0.1, 0.03)	0.05 (-0.01, 0.11)	0.02 (-0.09, 0.12)	-0.06 (-0.11, -0.01)	-0.03 (-0.08, 0.03)
Holdemanella	0.13 (0.05, 0.22)	0.00 (-0.09, 0.09)	-0.05 (-0.11, 0.01)	-0.07 (-0.13, -0.01)	-0.01 (-0.12, 0.09)	-0.03 (-0.08, 0.02)	-0.03 (-0.09, 0.02)
Lachnospira	0.01 (-0.04, 0.07)	-0.02 (-0.07, 0.03)	0.02 (-0.02, 0.06)	-0.03 (-0.07, 0.01)	0.06 (0.00, 0.12)	-0.01 (-0.04, 0.02)	0.00 (-0.03, 0.03)
Ruminococcaceae UCG-014	0.0 (-0.06, 0.07)	0.04 (-0.03, 0.1)	-0.03 (-0.08, 0.01)	-0.01 (-0.05, 0.03)	0.03 (-0.05, 0.1)	0.03 (-0.01, 0.06)	0.01 (-0.03, 0.05)
Coprococcus_2	-0.02 (-0.08, 0.04)	-0.01 (-0.08, 0.05)	-0.02 (-0.07, 0.02)	0.00 (-0.04, 0.04)	0.04 (-0.04, 0.11)	-0.02 (-0.05, 0.02)	-0.01 (-0.05, 0.03)
Clostridium_sensu_stricto_1	0.01 (-0.06, 0.08)	0.04 (-0.03, 0.12)	0.04 (-0.01, 0.09)	0.01 (-0.04, 0.06)	0.07 (-0.01, 0.15)	0.03 (-0.01, 0.07)	-0.01 (-0.05, 0.04)

Lachnospiraceae*	-0.08 (-0.15, 0.00)	-0.1 (-0.17, -0.03)	0.04 (-0.02, 0.09)	0.03 (-0.02, 0.08)	0.07 (-0.01, 0.16)	-0.06 (-0.1, -0.02)[1]	0.00 (-0.04, 0.05)
Faecalibacterium	-0.06 (-0.12, -0.01)	-0.05 (-0.1, 0.01)	-0.03 (-0.07, 0.01)	-0.01 (-0.05, 0.03)	0.06 (-0.01, 0.12)	-0.05 (-0.08, -0.02)[2]	-0.02 (-0.05, 0.02)
Lactobacillus	0.00 (-0.08, 0.09)	0.11 (0.02, 0.2)	-0.03 (-0.09, 0.03)	-0.02 (-0.08, 0.04)	0.02 (-0.09, 0.12)	0.01 (-0.04, 0.05)	-0.02 (-0.07, 0.04)
Subdoligranulum	-0.06 (-0.11, -0.02)	-0.01 (-0.06, 0.03)	-0.02 (-0.05, 0.02)	-0.03 (-0.06, 0.00)	0.02 (-0.04, 0.07)	-0.01 (-0.03, 0.01)	-0.02 (-0.04, 0.01)
Holdemanella	0.03 (-0.05, 0.12)	-0.13 (-0.22, -0.05)	-0.01 (-0.07, 0.05)	0.07 (0.01, 0.12)	0.05 (-0.05, 0.15)	-0.03 (-0.08, 0.01)	0.01 (-0.05, 0.06)
Lachnospiraceae_ ND3007_group	0.01 (-0.04,	-0.03 (-0.08,	-0.02 (-0.06,	-0.05 (-0.08,	0.03 (-0.03,	-0.04 (-0.06,	0.01 (-0.01,

	0.05)	0.01)	0.01)	-0.02)	0.09)	-0.01)[1]	0.04)
Fusicatenibacter	-0.04	0.02	0.01	0.01	0.01	-0.01	-0.01
	(-0.07,	(-0.01,	(-0.02,	(-0.01,	(-0.02,	(-0.03,	(-0.03,
	-0.01)	0.05)	0.03)	0.03)	0.05)	0.01)	0.01)
Subdoligranulum	-0.05	-0.01	0.04	-0.02	0.00	-0.02	-0.02
	(-0.09,	(-0.05,	(0.02,	(-0.05,	(-0.05,	(-0.04,	(-0.04,
	-0.01)	0.03)	0.07)	0.01)	0.04)	0.00)	0.01)
Akkermansia	0.03	0.11	-0.05	0.03	0.02	-0.03	-0.03
	(-0.05,	(0.03,	(-0.1,	(-0.02,	(-0.07,	(-0.07,	(-0.08,
	0.1)	0.19)	0.01)	0.08)	0.12)	0.01)	0.01)
Ruminococcaceae_	0.03	-0.08	0.04	0.04	0.02	0.02	-0.04
UCG-014	(-0.03,	(-0.15,	(0,	(-0.01,	(-0.06,	(-0.02,	(-0.08,
	0.1)	-0.01)	0.09)	0.08)	0.1)	0.05)	0.00)
Ruminococcaceae_	0.01	0.03	-0.01	0.02	0.02	-0.03	-0.03
UCG-004	(-0.04,	(-0.03,	(-0.05,	(-0.02,	(-0.05,	(-0.06,	(-0.06,
	0.07)	0.08)	0.03)	0.06)	0.08)	0.00)	0.01)
Eubacterium	0.05	0.06	0.02	-0.02	0.02	-0.05	-0.01
	(0.00,	(0.00,	(-0.02,	(-0.05,	(-0.04,	(-0.08,	(-0.04,

	0.1)	0.11)	0.06)	0.02)	0.08)	-0.02)²	0.02)
Family_XIII*	0.01	0.05	-0.02	0.00	0.02	0.01	0.00
	(-0.03,	(0.01,	(-0.05,	(-0.03,	(-0.03,	(-0.01,	(-0.02,
	0.05)	0.1)	0.01)	0.03)	0.07)	0.04)	0.03)
Clostridium_sensu_ stricto_1	0.08	-0.1	0.01	0.00	0.06	0.07	-0.06
	(0.01,	(-0.17,	(-0.04,	(-0.05,	(-0.02,	(0.03,	(-0.11,
	0.15)	-0.02)	0.07)	0.05)	0.15)	0.11)²	-0.01)
Christensenellaceae_ R-7_group	-0.02	-0.06	-0.01	0.01	0.01	-0.01	-0.02
	(-0.08,	(-0.13,	(-0.05,	(-0.04,	(-0.07,	(-0.05,	(-0.06,
	0.04)	0.00)	0.04)	0.05)	0.08)	0.02)	0.02)
Granulicatella	0.02	0.02	-0.01	0.01	-0.03	0.02	-0.02
	(-0.02,	(-0.02,	(-0.05,	(-0.02,	(-0.08,	(0.00,	(-0.05,
	0.07)	0.07)	0.02)	0.05)	0.03)	0.05)	0.01)
Lachnospiraceae*	-0.06	-0.02	0.01	0.02	0.06	-0.01	-0.01
	(-0.12,	(-0.09,	(-0.04,	(-0.02,	(-0.02,	(-0.05,	(-0.05,
	0.01)	0.05)	0.06)	0.07)	0.14)	0.02)	0.03)

Ruminococcus_1	0.00 (-0.05, 0.06)	0.06 (0, 0.12)	-0.03 (-0.07, 0.01)	0.01 (-0.03, 0.05)	0.02 (-0.05, 0.09)	-0.02 (-0.05, 0.01)	-0.02 (-0.05, 0.02)
Mogibacterium	0.00 (-0.06, 0.06)	-0.04 (-0.1, 0.02)	-0.02 (-0.06, 0.03)	-0.04 (-0.08, 0.00)	-0.05 (-0.12, 0.02)	0.00 (-0.04, 0.03)	-0.01 (-0.04, 0.03)

N=2,833. CI, confidence interval; "*", unknown genera from famly. Only OTUs with false discovery rate-adjusted p-value <0.10 for joint test of Na/K ratio and its interaction with province/megacity are shown here. OTUs were ordered by joint test q-values. OTU relative abundance was $\log_{10}$ transformed. Linear model was adjusted for age, sex, provinces or megacities, batch or plate run, urbanization, occupation, income, high school completion, energy intake, percent energy from animal food, fried food intake, physical activity, ever smoked, alcohol consumption, and use of probiotic, non-steroid anti-inflammatory drug, and proton pump inhibitor. Please see the Supplementary Table 6 in the supplemental files to see results for the full list of OTUs.

[1] false discovery rate-adjusted p-value <0.10;

[2] false discovery rate-adjusted p-value <0.05;

Table S4.8. Province-speicifc estimates [coefficient (95% CI)] for the associations between sodium density, potassium density, and sodium-to-potassium (Na/K ratio) with individual metabolites

Metabolites	**Class**	**Pathway**	**Guizhou**	**Hunan**
		Sodium density		
N6-methyladenosine	Nucleotide	Purine Metabolism, Adenine containing	0.1 (-0.1, 0.3)	-0.19 (-0.28, -0.09)
1,2,3-benzenetriol sulfate (2)	Xenobiotics	Chemical	-1.21 (-1.82, -0.6)[2]	0.16 (-0.14, 0.46)
3-methoxycatechol sulfate (1)	Xenobiotics	Benzoate Metabolism	-1.08 (-1.6, -0.56)[2]	0.07 (-0.18, 0.32)
4-methylcatechol sulfate	Xenobiotics	Benzoate Metabolism	-0.73 (-1.08, -0.38)[2]	-0.07 (-0.24, 0.1)
		Potassium density		
6-oxopiperidine-2-carboxylate	Amino Acid	Lysine Metabolism	0.26 (-0.32, 0.85)	0.95 (0.53, 1.37)[2]
		Na/K ratio		
DSGEGDFXAEGGGVR*	Peptide	Fibrinogen Cleavage Peptide	0.23 (0.09, 0.38)	-0.15 (-0.24, -0.06)
N6-methyladenosine	Nucleotide	Purine Metabolism, Adenine containing	0.12 (0.02, 0.23)	-0.13 (-0.2, -0.07)
glycerophosphorylcholine (GPC)	Lipid	Phospholipid Metabolism	0.07 (0.04, 0.1)[2]	0 (-0.02, 0.02)
5-HETrE	Lipid	Eicosanoid	0.33 (0.16, 0.5)[1]	-0.09 (-0.19, 0.02)
4-HDoHE	Lipid	Docosanoid	0.2 (0.05, 0.34)	-0.14 (-0.23, -0.05)
4-guanidinobutanoate	Amino Acid	Guanidino and Acetamido Metabolism	0.1 (0.01, 0.19)	-0.08 (-0.14, -0.03)
thyroxine	Amino Acid	Tyrosine Metabolism	0.18 (0.09, 0.28)[1]	0.01 (-0.05, 0.07)

5-HETE	Lipid	Eicosanoid	0.24 (0.08, 0.39)	-0.1 (-0.2, -0.01)
13-HODE + 9-HODE	Lipid	Fatty Acid, Monohydroxy	0.16 (0.04, 0.28)	-0.09 (-0.17, -0.02)
isovalerate (C5)	Amino Acid	Leucine, Isoleucine and Valine Metabolism	0.16 (0.06, 0.26)	-0.07 (-0.13, 0.00)
5-HEPE	Lipid	Eicosanoid	0.28 (0.1, 0.46)	-0.1 (-0.22, 0.01)
butyrate/isobutyrate (4:0)	Lipid	Short Chain Fatty Acid	0.13 (0.05, 0.22)	-0.05 (-0.1, 0.01)
gamma-glutamyl-epsilon-lysine	Peptide	Gamma-glutamyl Amino Acid	0.08 (-0.02, 0.18)	-0.11 (-0.17, -0.04)
pipecolate	Amino Acid	Lysine Metabolism	-0.2 (-0.31, -0.1)[1]	-0.02 (-0.09, 0.05)

N=392. CI, confidence interval; Only metabolites with false discovery rate-adjusted p-value <0.10 for joint test of sodium density, potassium density, or Na/K ratio and its interaction with province/megacity are shown here. Metabolites were ordered by joint test q-values. Metabolite abundance was $\log_2$ transformed. Linear model was adjusted for age, sex, provinces or megacities, batch run, urbanization, occupation, income, high school completion, energy intake, percent energy from animal food, fried food intake, physical activity, ever smoked, alcohol consumption, and use of probiotic, non-steroid anti-inflammatory drug, and proton pump inhibitor. Please see the Supplementary Table 7-9 in the supplemental files to see results for the full list of metabolites.

[1] false discovery rate-adjusted p-value <0.10;

[2] false discovery rate-adjusted p-value <0.05;

Table S4.9. The associations between sodium density-, potassium density-, or sodium-to-potassium (Na/K) ratio-associated taxa and metabolites

Taxon	**Metabolite**	**Class**	**Pathway**	**Coefficient (95% CI)**	**q-value**
Coriobacteriaceae*	4-methylcatechol sulfate	Xenobiotics	Benzoate Metabolism	0.93 (0.46, 1.40)	0.08
Ruminococcaceae*	4-methylcatechol sulfate	Xenobiotics	Benzoate Metabolism	1.16 (0.57, 1.74)	0.08

N=392. "*", unknown genera from family. CI, confidence interval; q-value, false discovery rate-adjusted p-value; Only taxon-metabolite pairs q-v <0.10 are shown here. Taxon relative abudance was $\log_{10}$ transformed and metabolite abundance was $\log_2$ transformed. Linear model was adjusted for age, sex, provinces or megacities, plate or batch run, urbanization, occupation, income, high school completion, energy intake, percent energy from animal food, fried food intake, physical activity, ever smoked, alcohol consumption, and use of probiotic, non-steroid anti-inflammatory drug, and proton pump inhibitor. Please see the Supplementary Table 10 in the supplemental files to see results for the full list of taxon-metabolite pairs.

Figure S4.1. Sample flow chart.

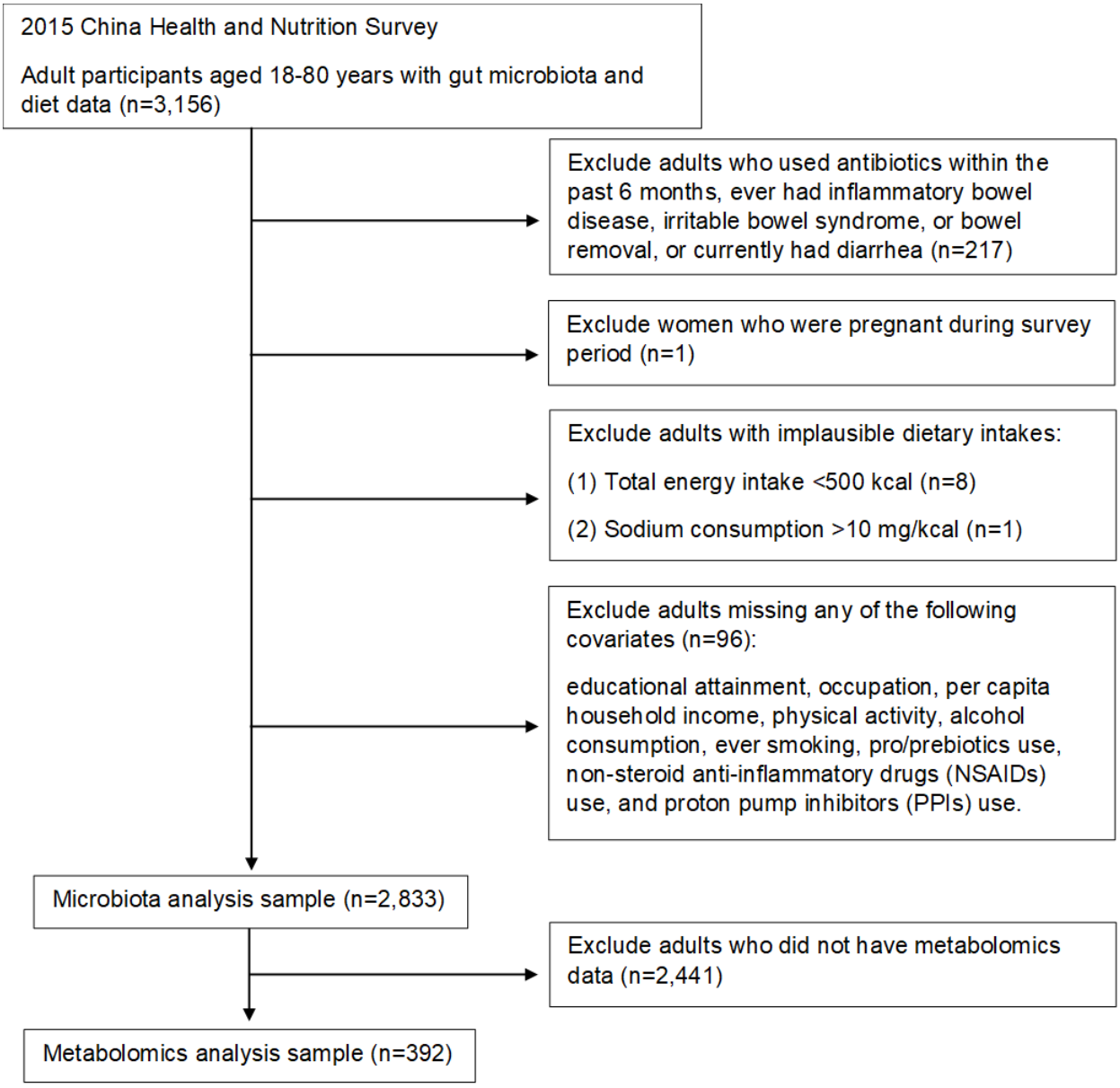


Figure S4.2. Multidimensional scaling (MDS) plots of gut microbiota with respect to sodium density (A) across and (B) within provinces and megacities. Numbers within parentheses are percentage of variability in microbial similarity explained by MDS axes.

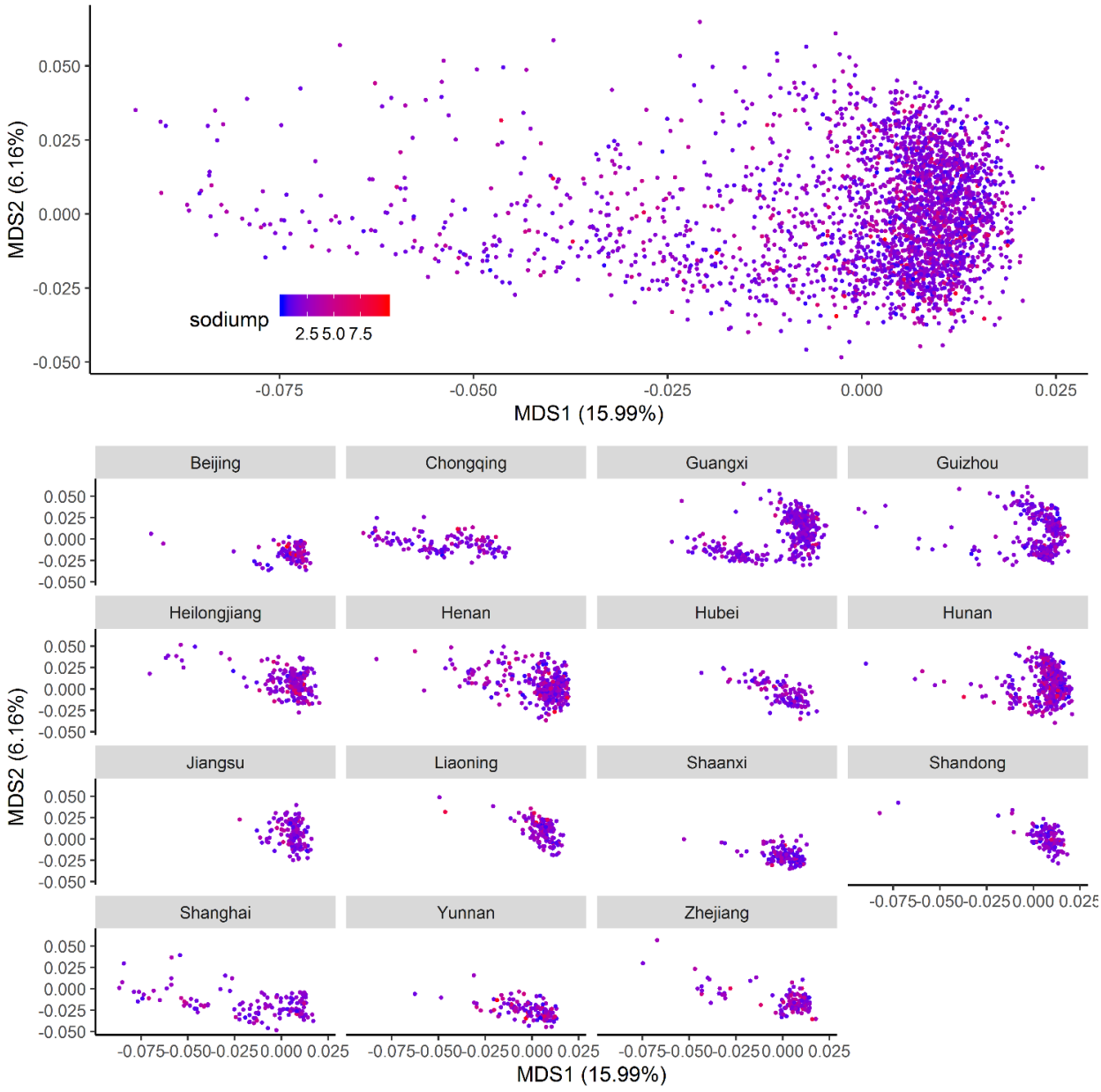


Figure S4.3. Multidimensional scaling (MDS) plots of gut microbiota with respect to potassium density (A) across and (B) within provinces and megacities. Numbers within parentheses are percentage of variability in microbial similarity explained by MDS axes.

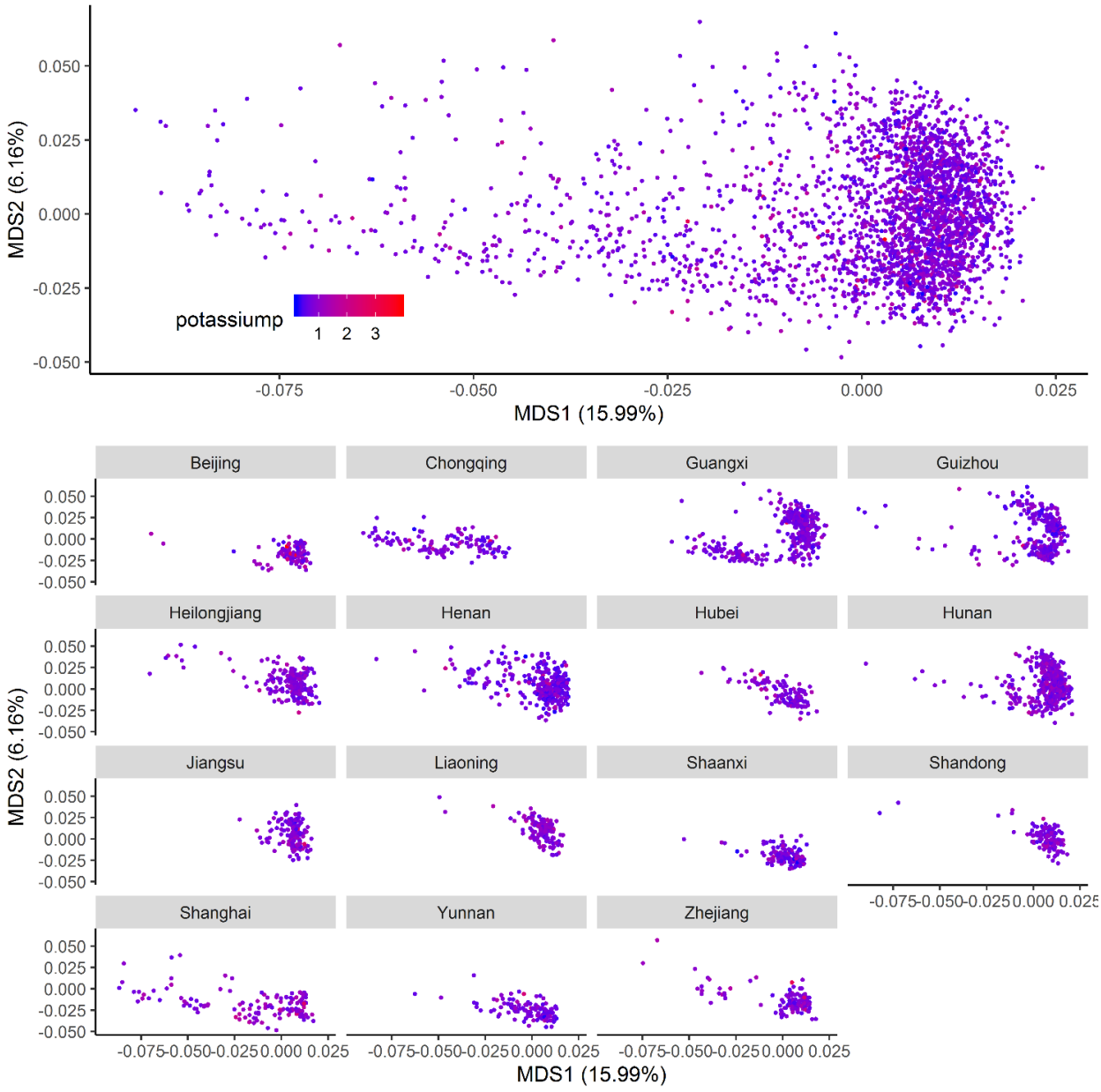


Figure S4.4. Multidimensional scaling (MDS) plots of gut microbiota with respect to sodium-to-potassium (Na/K) ratio (A) across and (B) within provinces and megacities. Numbers within parentheses are percentage of variability in microbial similarity explained by MDS axes.

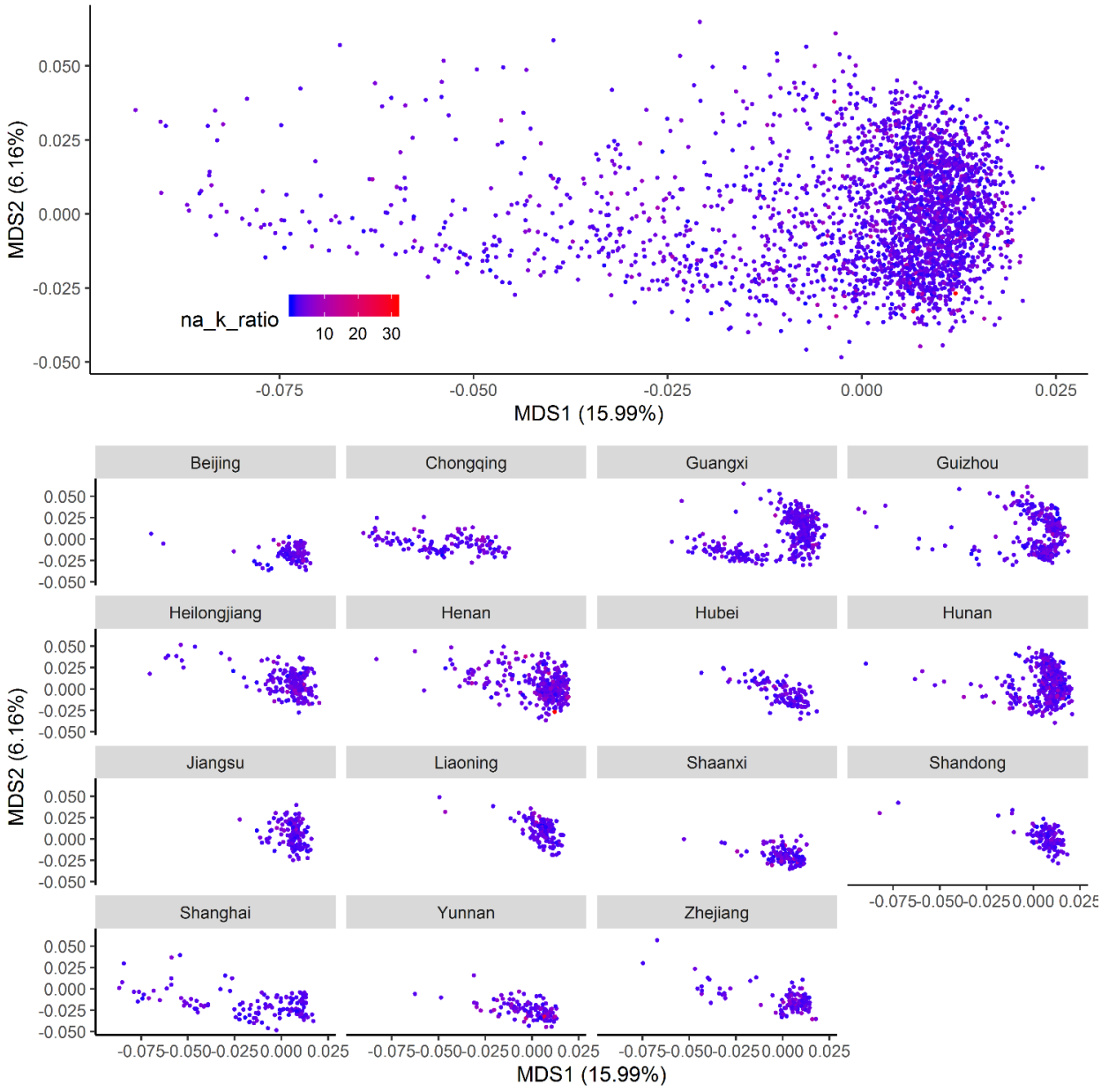


Figure S4.5. Multidimensional scaling (MDS) plots of plasma metabolome with respect to sodium density (A) across and (B) within provinces. Numbers within parentheses are percentage of variability in microbial similarity explained by MDS axes.

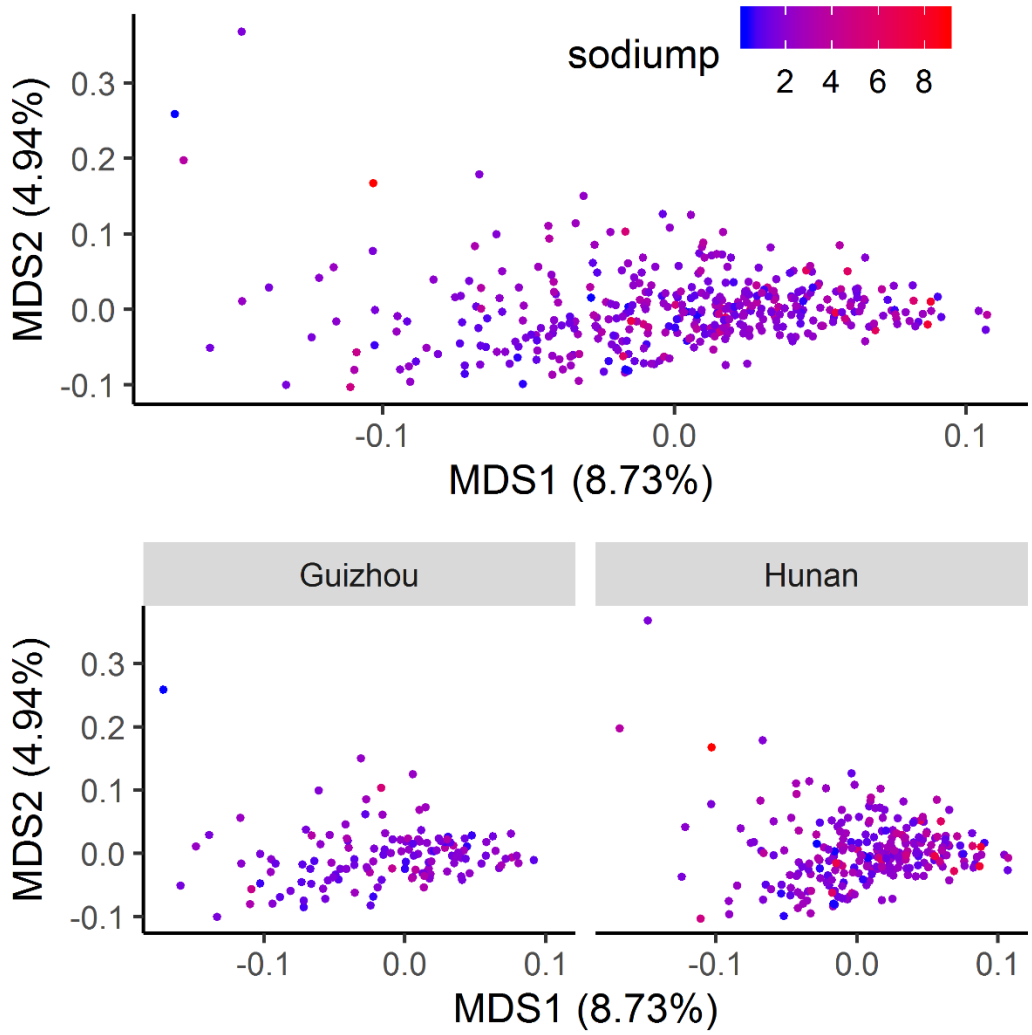


Figure S4.6. Multidimensional scaling (MDS) plots of plasma metabolome with respect to potassium density (A) across and (B) within provinces. Numbers within parentheses are percentage of variability in microbial similarity explained by MDS axes.

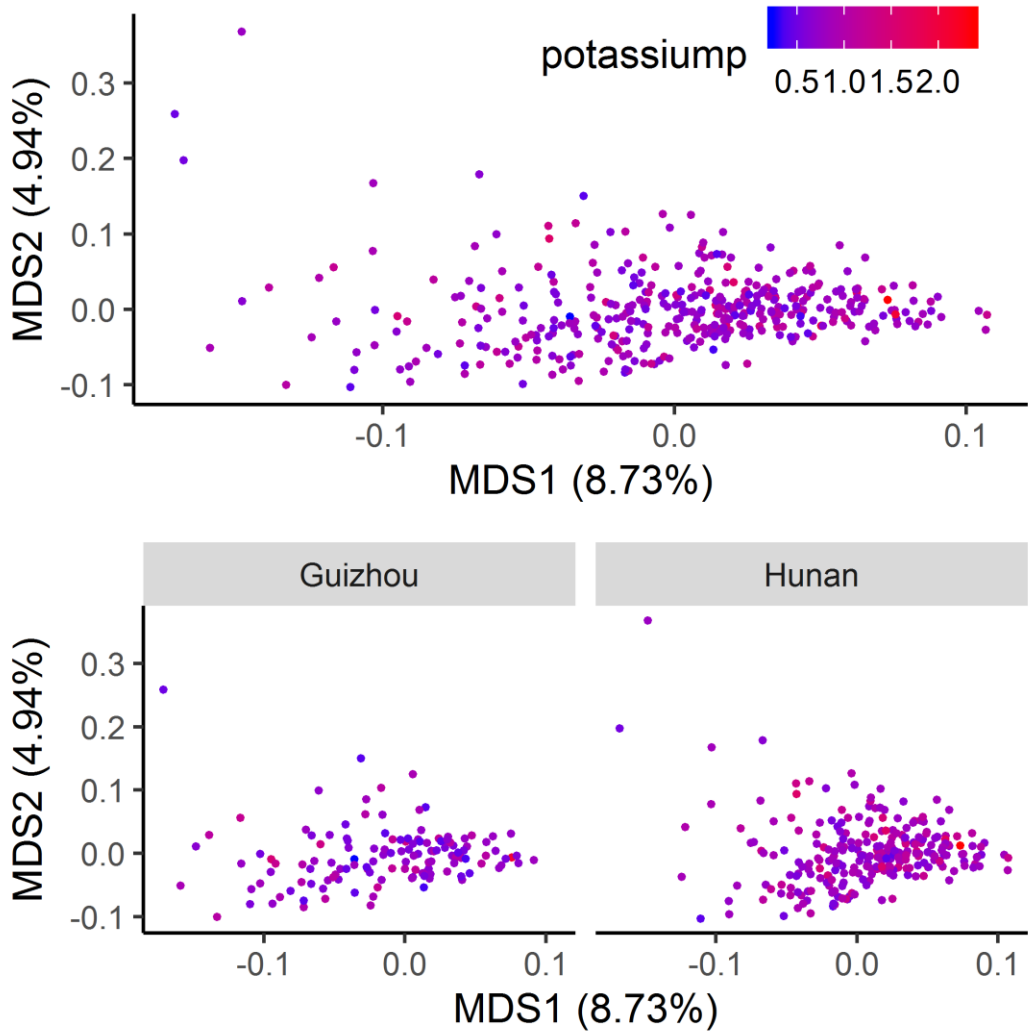
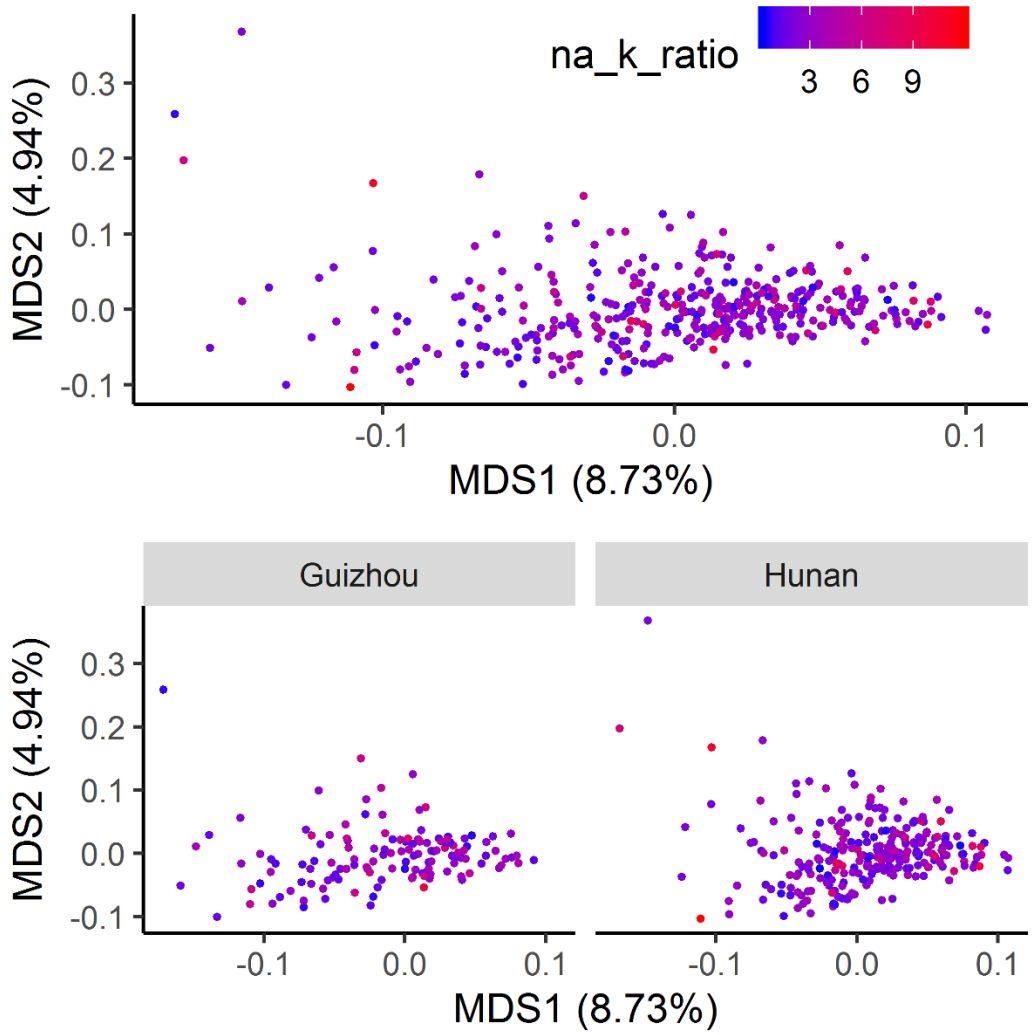


Figure S4.7. Multidimensional scaling (MDS) plots of plasma metabolome with respect to sodium-to-potassium (Na/K) ratio (A) across and (B) within provinces. Numbers within parentheses are percentage of variability in microbial similarity explained by MDS axes.



CHAPTER 5. GUT MICROBIOTA AND HOST METABOLITES ASSOCIATIONS WITH BLOOD PRESSURE IN CHINESE ADULTS

Overview

Animal studies have revealed gut microbial and metabolic pathways of blood pressure (BP) regulation, yet there are few epidemiological studies that capture large variation in BP with paired microbiota and metabolomics data.

In a population-based, Chinese cohort (30-69 years, 54% women), we examined cross-sectional associations of gut microbiota (16S rRNA, n=1003) and untargeted plasma metabolomics (n=434) with systolic and diastolic BP (SBP/DBP), after adjusting for a wide range of covariates (e.g., urbanization, diet, kidney function). We found that the overall microbial community assessed by principal coordinates analysis based on Bray-Curtis matrix varied by SBP and DBP (permutational multivariate ANOVA p-value<0.05). No specific genera were associated with SBP or DBP using linear regression.

In metabolomics analysis, a lipid pattern derived from principal component analysis was positively associated with SBP [linear regression coefficient (95% CI) per 1SD pattern score: 2.23 (0.72, 3.74) mmHg] and DBP [1.72 (0.81, 2.63) mmHg]. Individual metabolites including linoleate, palmitate, and dihomolinolenate, as well as eight sphingomyelins, four acyl-carnitines, and two phosphatidylinositol, were positively associated with SBP and DBP [false discovery rate (FDR) adjusted linear model p-value<0.05). Subsequent pathway analysis suggested that metabolites from acyl-carnitine (long chain saturated), phosphatidylinositol, and sphingomyelins metabolic pathways were positively associated with SBP and DBP more than expected by chance (FDR adjusted Fisher's exact test p-value<0.05).

Our results suggest potential metabolic pathways involved in BP regulation in free-living adults, to be followed up in future human intervention and clinical studies.

Introduction

High blood pressure (BP) is a leading modifiable risk factor for cardiovascular disease and mortality [2]. Despite numerous efforts to curb the epidemic, the worldwide prevalence of high BP has continued to increase over the past decade [16] and the prevalence of controlled high BP has remained low [140]. In the US, only 48.3% of adults with high BP reduced their systolic and diastolic BP (SBP/DBP) below the threshold for high BP (140/90 mmHg) after treatment during 2015-2016 [140].

The BP regulatory system is multifactorial, involving interactions among host genetics [141], sociodemographic factors, and diet [24]. The gut microbiota and host metabolome, which may reflect these complex interactions [106,142], have been demonstrated to play fundamental roles in BP regulation in animal models [80,143] and humans [33]. In particular, the metabolome reflects a thorough snapshot of various metabolic processes, allowing identification of novel biomarkers and potential pathogenic pathways leading to high BP [144]. For example, the microbiota-mediated serum 4-hydroxyhippurate has been shown to be positively associated with incident high BP in blacks [49]. Additionally, reductions in gut microbial diversity and in several microbial groups, including *Prevotella* and *Coprococcus*, have been shown to be associated with high BP in animal models [36,37,40] and in human studies, albeit in small samples [39,41,43,145]. However, there is a lack of population-based studies that include microbial and metabolomic data along with phenotypic data. In comparison with the rest of the world, China has the greatest absolute burden of high BP [13] coupled with a high rates of undiagnosed and untreated hypertension [146], thus China is an ideal context for studying relationships between BP, microbiota, and metabolites, while minimizing the medication effects.

To this end, we examined the associations between gut microbiota and nontargeted plasma metabolome with BP in a well-characterized cohort of adults from the 2015 China Health and Nutrition

Survey (CHNS). We quantified the associations between specific microbiota and metabolites with BP, and identified patterning of microbiota and metabolites associated with BP.

Methods

Study sample

We used data from the 2015 China Health and Nutrition Survey (CHNS). Participants aged 30-68 years from four southern provinces (Henan, Hunan, Guizhou, Guangxi) with blood pressure data and gut microbiome or plasma metabolomics data were eligible for analysis (n=1,285, Figure S5.1). We excluded participants who were pregnant (n=1), self-reported use of antihypertension medication (n=99), or had missing covariates (n=86). For microbiota analysis, we additionally excluded 35 participants who currently used antibiotics, had diarrhea, inflammatory bowel disease, irritable bowel syndrome, or bowel removal. For metabolites analysis, we additionally excluded 16 participants who had detectable levels of CVD dugs metabolites in plasma: metoprolol acid metabolite, alpha-hydroxymetoprolol, nifedipine, and valsartan. The total analysis sample had 1,082 adults, with 1,003 and 434 adults included in the microbiota analysis sample and metabolomics analysis sample, respectively.

Measures

We used the average of the three readings of reasting BP measured by experienced physicians using a standard mercury sphygmomanometer as our measure of SBP and DBP.

In 16S rRNA sequencing, a total of 1027 genera were detected in our sample, 19 of which with all zero values were removed from analysis. We rarefied the resulting raw taxonomic counts to 21,600 sequences/sample to correct for different sequencing depth (before rarefaction, 21,648-89,427 sequences/sample) and $\log_{10}$ transformed it.

The non-targeted metabolomics analysis of fasting plasma samples using Metabolon portal resulted in the detection and quantification of 1,108 chemicals in our sample, among which four CVD drugs with all zero values were removed from analysis. We categorized 131 metabolites that below

detection limits (BDL) in 25%-50% samples to ordinal variables (0=BDL, 1=below median, 3=equal or above median) and 99 metabolites >50% BDL to binary variables (0=BDL, 1=above detection limit). For 874 metabolites with ≤25% BDL, we used the rescaled-imputed data from Metabolon (raw area counts of each metabolite rescaled to a median of 1 with values BDL imputed by the minimum value) and $\log_2$ transformed the data.

We included the following measures as covariates in our analysese based on self-reported sociodemographic and behavioral information collected using standardized questionnaires administered by trained interviewers, such as age, sex, education (yes/no completed high school), and per capita household income (household income / number of household member). We assessed community-level urbanization using community-level data and a validated urbanization index that encompasses 12 dimensions of urbanization [109]. We included five measures of heath behaviors: the total energy intake, sodium intake, total physical activity (METs/week), ever smoking (yes/no), and alcohol intake in the past year (yes/no). We also included a measure of kidney function based on fasting serum creatinine concentration measured by picric acid method using Hitachi 7600 (Randox, UK) for calculation of estimated glomerular filtration rate (eGFR) using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation [147]. We included body mass index (BMI) as a measure of body mass.

Statistical analysis

Primary outcomes were SBP and DBP. In the microbiota analysis sample, we first analyzed the overall gut microbiome by examining the associations of genus-level within-person microbial diversity (α-diversity), measured by Shannon index and richness [116,148], and between-person diversity (β-diversity), assessed by principal coordinate analysis (PCoA) based on Bray-Curtis dissimilarity matrix [149], with SBP and DBP using linear regression and permutational multivariate analysis of variance (PERMANOVA) with 999 permutations [150], respectively. PCoA axis score was a weighted (i.e., genera relative abundance) sum of genera scores, which reflect the contribution of each genera to a given

axis (Table S5.1). Then, we quantified the association between each of the first four PCoA axes with SBP and DBP. Second, we examined the association between specific microbiota across 1,008 specific genera, with SBP and DBP using linear regression. We treated 110 genera detected in ≥25% of the sample as continuous variables and dichotomized the rest 898 rare genera to presence/absence. We adjusted all analyses for the following potential confounders in Model 1 using *a priori* knowledge: age, sex, provinces [151], urbanization index (tertiles) [119], education, per-capita household income (tertiles), total energy intake, animal-source food consumption [120], sodium consumption [152], total physical activity (tertiles), tobacco use, alcohol consumption, and eGFR [153]. As BMI is a potential mediator for microbiota-BP relationship, we additionally adjusted for BMI in Model 2 as a sensitivity analysis to test whether the association was independent of BMI.

In the metabolomics analysis sample, we first analyzed the overall metabolome by separately grouping 875 metabolites (continuous variables, ≤25% BDL) into uncorrelated patterns, using principal component analysis (PCA) followed by a varimax rotation to improve interpretation [154] and account for complex correlations across metabolites. Based on three criteria: eigenvalues >1, the point of inflection in scree plot, and interpretability [155], we selected three metabolite patterns (Table S5.2). Pattern score was a weighted (i.e., metabolite relative abundance) sum of rotated and inverse factor loadings, which indicate the contributions of metabolites to a given pattern. Then, we assessed the association between each metabolite pattern, as well as 1,104 individual metabolites with SBP and DBP, using linear regression adjusting for the above-mentioned potential covariates (e.g., province, sodium consumption, eGFR) and batch in Model 1 and additionally adjusting for BMI in Model 2. We used a Wald test to assess the overall statistical significance of 131 metabolites categorized as ordinal variables. Based on Model 1 results for individual metabolites, we calculated pathway enrichment score $[\frac{k}{m}/(\frac{n-k}{N-m})]$ reflecting the degree to which a given pathway was associated with SBP or DBP, where k and n are numbers of statistically significant metabolites in the given pathway and all pathways, respectively, and m and N are numbers of tested metabolites in the given pathway and all pathways, respectively. We performed a

Fisher's exact test [156] to evaluate whether the presence of blood pressure-associated metabolites (excluding unknown compounds) from a particular metabolic pathway was greater than expected by chance.

In a sub-sample of participants with microbiota and metabolite data (n=355), we conducted random forest regression (100 trees) [125] followed by a 5 iterations of 2-folds cross-validation (5X2cv) modified paired t-test, a powerful test to compare the performance of learning algorithms with acceptable Type I error [126], to provide insight into which data as a whole had the strongest association with BP. Specifically, we compared the accuracy (i.e., root mean squared errors, RMSE) of prediction of SBP and DBP using: (1) host factors (e.g.,14 individual- and community-level covariates), (2) microbiota (e.g., 1,008 specific genera), (3) metabolites (e.g.,1,104 individual metabolites), (4) microbiota + host factors, (5) metabolites + host factors, (6) microbiota + metabolites, and (7) microbiota + metabolites + host factors.

We adjusted p-values for multiple comparisons using Benjamini-Hochberg method (false discovery rate, FDR) [127] in comparisons across all taxa, metabolites, and metabolic pathways for SBP and DBP separately in a test of two separate hypotheses for SBP and DBP, respectively. All statistical tests were two-sided with a significance level of 0.05. We used R 3.6.0 (http://www.r-project.org) and Python 3.5.1 (https://www.python.org) for data analysis.

Results

Our sample had large variation in SBP [mean (SD): 126.01 (17.43) mmHg] and DBP [80.70 (10.67)], with 62.48% prevalence of high blood pressure (Table S5.3).

We first assessed the overall gut microbial measures. We found that while within-person microbial diversity (Shannon index and richness) was not associated with SBP or DBP in linear regression (Table S5.4), between-person microbial diversity assessed by PCoA was associated with SBP (Figure 5.1; PERMANOVA R^2=2.01%, p-value=0.002) and DBP (Figure S5.2; PERMANOVA R^2=1.42%, p-value<0.05). The first four PCoA axes each explained 8.61%, 5.58%, 3.54%, and 3.12% of

microbial variability, respectively. Only the fourth axis showed a clear separation of SBP (Fugure 5.1, with higher SBP at higher axis score. While *Rothia, Serratia, Enterobacteriaceae, Leuconostocaceae,* and *Fusobacterium* had the strongest positive correlations with the fourth axis, *Coprococcus, Adlercreutzia, Eggerthella,* and *Raistonia* had the strongest negative correlations with the fourth axis. The linear regression also showed a positive association between the fourth PCoA axis and SBP independent of BMI (Table S5.5). However, none of the specific genera were associated with SBP or DBP (Table S5.6, FDR-adjusted p-valueq-value>0.10).

In plasma metabolomics analysis, we identified three biologically possible metabolite patterns using PCA that explained a total of 19.11% variance of metabolites (Table 5.1). Only the second pattern that characterized by lipids like linoleate, palmitate, and oleate/vaccinate was positively associated with SBP [linear model coefficient (95% CI) per 1SD pattern score: 2.23 (0.72, 3.74)] and DBP [1.72 (0.81, 2.63)], though the results were slightly attenuated by adjustment of BMI [SBP: 1.88 (0.38, 3.38); DBP: 1.45 (0.55, 2.35)]. To assess whether single or a few metabolites contributing to this lipid pattern drove the associations with SBP and DBP, we examined individual metabolites using linear regression and found that at Model 1 FDR-adjusted p-value<0.05, eight and 19 metabolites of the lipid pattern (loading>0.4) were positively associated with SBP (Table 5.2) and DBP (Table 5.3), respectively, including palmitoylcarnitine (C16), cerotoylcarnitine (C26), 1-palmitoleoylglycerol (16:1), myristoylcarnitine (C14), dihomolinolenate (20:3n3 or 3n6), laurylcarnitine (C12). Among all detected metabolites, a total of 34 and 39 metabolites were associated with SBP and DBP, respectively, among which 19 metabolites were positively associated with both BP measures, including eight sphingomyelins [e.g., tricosanoyl sphingomyelin (d18:1/23:0), lignoceroyl sphingomyelin (d18:1/24:0)], four acyl-carnitines, and cholesterol. Only nine and 17 metabolites were associated with SBP and DBP independent of BMI (Model 2 FDR-adjusted p-value<0.05), respectively, including tricosanoyl sphingomyelin (d18:1/23:0), lignoceroyl sphingomyelin (d18:1/24:0), and sphingomyelin (d18:2/24:2). A full list of results for all individual metabolites are shown in Supplementary Table S12-S14 in supplemental file. In pathway analysis that tested whether the number of positive or negative associations between metabolites

from a particular metabolic pathway and BP was more than expected by chance (Table 5.4), we found that diacylglycerol, acyl-carnitine (long chain saturated), phosphatidylcholine, phosphatidylinositol, sphingomyelins metabolic pathways were associated with SBP (FDR-adjusted p-value<0.05); and corticosteroids, acyl-carnitine (long chain saturated and median chain), monoacylglycerol, phosphatidylinositol, and sphingomyelins metabolic pathways were associated with DBP.

To examine whether the overall microbiota and metabolite data were better than traditional risk factors like sociodemographic factors and health behaviors in predicting BP, we conducted random forest regression in a sub-sample with microbiota and metabolite data (n=355). We found comparable accuracies across host factors, microbiota, and metabolite data in predicting SBP and DBP (Figure S5.3, p-value>0.05).

Discussion

We investigated the association between gut microbiota and plasma metabolites with BP in a population-based cohort of middle-aged Chinese adults, after accounting for a wide range of sociodemographic factors, health behaviors, and kidney function. Between-person diversity in microbiota was associated with SBP and DBP, whereas there was no statistical evidence of any association between within-person diversity or with specific genera in relation to SBP or DBP. In metabolomics analysis, a lipid pattern that included various long-chain fatty acids like linoleate, palmitate, and oleate/vaccenate was positively associated with SBP and DBP, independent of BMI. Several individual metabolites were also associated with SBP (n=34) and DBP (n=39), including eight sphingomyelins, four acyl-carnitines, and cholesterol, which were positively associated with SBP and DBP.

Several animal [36,37,40] and human studies [39,41,43,145] have linked gut microbiota to high blood pressure. For example, in the CARDIA study of 529 middle-aged US adults, Sun et al. [43] showed an inverse association between within-person microbial diversity and SBP and differences in the overall microbial community by SBP. In a recent case-control study of 80 Brazilian adults [145], lower microbial biodiversity along with lower proportions of butyrate-producing taxa like *Roseburia, Coprococcus* and

Lachnospiraceae, but higher proportions of *Enterobacteriaceae* and *Lactobacillus* were observed in individuals with high versus normal blood pressure. Similarly, we found an overall association between the microbial community with SBP and DBP. Although we had larger sample size than these studies [43,145], we found no statistical evidence of any association between specific genera with blood pressure. As such, our results may suggest that instead of a few specific taxa, it is the overall microbiota composition or a collection of microorganisms that associated with blood pressure. In addition, by excluding all participants who took antihypertension medication, our study may have better control for potential hypertension treatment effects on gut microbiota than prior studies [43,145].

Metabolomics studies showing associations between microbial metabolites and blood pressure further support the role of gut microbiota in blood pressure regulation [49,157]. The INTERMAP study (International Population Study on Macronutrients and Blood Pressure) of 4,630 middle-aged adults from USA, UK, Japan, and China, showed that urinary alanine and hippurate were positively and negatively associated with blood pressure, respectively [157]. The Atherosclerosis Risk in Communities (ARIC) study of 896 African Americans revealed that 4-hydroxyhippurate was associated with 17% higher risk of incident high blood pressure [49]. Though we did not observe an associations between alanine or hippurate metabolites with blood pressure in our sample, we found that a novel metabolite from benzoate metabolism, p-cresol sulphate, which is produced from tyrosine and phenylalanine by anaerobic bacteria [158], was inversely associated with DBP. A well-establish route through which gut microbiota influence blood pressure is short chain fatty acids (SCFAs), like butyrate and propionate, which have been shown to modulate blood pressure through G protein-coupled receptors and olfactory receptors [79]. Nonetheless, none of these studies (INTERMAP and ARIC) [49,157] nor our study found an association between SCFAs and blood pressure. This may be due to the small molecular sizes and rapid uptake of SCFAs in circulation [159], impeding them from being fully captured by non-targeted metabolomics.

Prior studies also suggested that mechanisms of blood pressure regulation involve fatty acids [49,144]. Menni et al. showed that in 3,980 TwinsUK females, a few plasma carnitines, long chain fatty acids, and steroids were positively associated with blood pressure, including hexadecanedioate, palmitate

(16:0), octanoylcarnitine, 10-heptadecenoate (17:1n7), and dihomolinoleate (20:2n6) [144]. In particular, hexadecanedioate, a dicarboxylic acid, consistently showed positive association with blood pressure in two replication cohorts with both males and females, the Cooperative Health Research in the Region of Augsburg (KORA) S4 study and Hertfordshire Cohort [144]. Subsequent analysis using rat model demonstrated that oral intake of hexadecanedioate increased blood pressure, supporting a causal role of hexadecanedioate in blood pressure regulation [144]. In another study of 202 African and Caucasian men, serum long-chain [e.g., cerotoylcarnitine (C26)] and medium-chain acyl-carnitines [in Caucasian only; e.g., octanoylcarnitine (C8)] were positively associated with ambulatory blood pressure.[160] Our results showing positive associations between acyl-carnitines (e.g., octanoylcarnitine), long chain fatty acids (e.g., palmitate), and a lipid pattern driven by linoleate (18:2n6), palmitate (16:0), 10-heptadecenoate (17:1n7), and dihomolinoleate (20:2n6) with BP are consistent with these two studies [144,160]. Acyl-carnitines as byproducts of incomplete β-oxidation of long-chain fatty acids accumulate in blood or urine when there are excess fatty acids for oxidation and can stimulate proinflammatory pathways involving Nuclear factor kappa B (NF-κB) [161]. Omega-6 fatty acid like linoleate may impair cardiovascular health as it can be metabolized to dihomolinoleate and then to arachidonic acid, a precursor for proinflammatory eicosanoids like leukotriene B4 [162].

In addition, we found that several sphingomyelins and the sphingomyelin metabolic pathway were each positively associated with blood pressure. Ceramide as a precursor for sphingolipids is deleterious to cardiovascular health, including impaired vasodilation [163]. Excess sphingolipids occur when fatty acids exceed energy need or storage capacity of a cell [164]. Several lipidomic studies have identified sphingolipids as candidate blood markers for cardiovascular diseases in humans [165-167]. For example, Poss et al. [167] using machine learning found that 30 serum sphingolipids were elevated in subjects with coronary artery disease (CAD, n=462) than controls (n=212) and a sphingolipid risk score, which included sphingomyelins (d18:1/24:0) and (d18:1/18:0), was more effective than conventional biomarkers like triglycerides and LDL-cholesterol in distinguishing CAD patients. Overall, our results suggest that overloaded circulating lipids are associated with higher blood pressure.

The strengths of our study include paired microbiota and metabolite data in a well-characterized cohort, enabling us to investigate microbial and host metabolic pathways in relation to BP. Fecal samples and blood samples were collected and processed using standardized protocols. BP was measured by trained clinicians using standardized protocols as well. Moreover, the rich sociodemographic and behavioral data of the CHNS allowed us to account for a wide range of potential confounders, including diet assessed by validated instruments, as well as anthropometry and eGFR measured by clinicians. The low treatment rate for high BP ensured sufficient sample size and large variation in BP, even after excluding people who took antihypertension medication to minimize medication effects.

However, our study is cross-sectional, which hinders the establishment of a causal relationship between gut microbiota, host metabolome, and blood pressure. Additionally, our microbiota analysis using 16S rRNA data did provide functional information and thus could not directly link to our results in metabolites. Future studies are needed to confirm our findings, particularly, population-based studies with repeated measures paired with experimental studies to investigate the causal pathways modulating blood pressure.

Conclusion

Our study provides substantial observational evidence for the associations between gut microbiota and plasma metabolites with BP in a population-based cohort of middle-aged Chinese adults. The overall microbial community varied by BP. Several individual metabolites (e.g., lignoceroyl sphingomyelin, cerotoylcarnitine, and dihomolinolenate) and a lipids pattern were positively associated with BP, suggesting a role of circulating lipids in the development of high BP. Further analyses with longitudinal data and refined microbial composition data in larger samples are needed to fully elucidate the causal relationship between gut microbiota, host metabolites, and BP, thereby informing effective early interventions and treatments for high BP.

Tables and figures

Table 5.1. Association between metabolite patterns with systolic and diastolic blood pressure (SBP and DBP, mmHg), Coefficient (95% CI)

	Metabolites contributing to each pattern		**Eigenvalue**	**Variance explained**	**SBP**		**DBP**	
					Model 1	**Model 2**	**Model 1**	**Model 2**
Pattern 1 (nucleotide, amino acid, and peptide)	pseudouridine; 2,3-dihydroxy-5-methylthio-4-pentenoate (DMTPA); N-acetylthreonine; N,N-dimethyl-pro-pro; C-glycosyltryptophan;	orotidine; hydroxy-N6,N6,N6-trimethyllysine; 5,6-dihydrouridine; dimethylarginine (ADMA + SDMA); N6-acetyllysine	84.19	9.63%	1.81 (-0.24, 3.86)	1.58 (-0.44, 3.60)	0.24 (-1.01, 1.49)	0.06 (-1.16, 1.28)
Pattern 2 (lipids, especially long chain fatty acids)	linoleate (18:2n6); palmitate (16:0); oleate/vaccenate (18:1); 10-heptadecenoate (17:1n7);	10-nonadecenoate (19:1n9); margarate (17:0); dihomolinoleate (20:2n6);	41.83	4.79%	2.23 (0.72, 3.74) **	1.88 (0.38, 3.38) *	1.72 (0.81, 2.63) **	1.45 (0.55, 2.35) **

	docosapentaenoate (DPA; 22:5n3); hexadecadienoate (16:2n6);	dihomolinolenate (20:3n3 or 3n6)						
Pattern 3 (sphingomyelins, eicosanoid, and short-chain fatty acids, and branched-chain amino acids)	sphingomyelin (d18:2/23:0, d18:1/23:1, d17:1/24:1); 3-methyl-2-oxobutyrate; leukotriene B4; 5-HETE; methionine sulfoxide; butyrate/isobutyrate (4:0);	4-methyl-2-oxopentanoate; 1-(1-enyl-oleoyl)-GPE (P-18:1); 1-(1-enyl-palmitoyl)-GPE (P-16:0); 3-methyl-2-oxovalerate	40.99	4.69%	0.24 (-1.57, 2.05)	-0.02 (-1.80, 1.77)	0.14 (-0.96, 1.24)	-0.05 (-1.13, 1.02)

Patterns were derived from principal component analysis followed by a varimax rotation of 874 metabolites in the metabolomics analysis sample (n=434). The factor loading for each metabolite contributing to each pattern is listed in Table S5.2. Coefficient (95% confidence interval) indicates SBP and DBP (mmHg) associated with each 1SD increase in metabolites pattern score in linear regression. Contributing metabolites are metabolites with the top 10 highest absolute loadings for the respective pattern. Metabolites are vertically ordered by their absolute values of loading (descending). Model 1 was adjusted for age, sex, provinces, batch, urbanization index (tertiles: ≤64.2, 64.2-81.5, >81.5), per capita household income (tertiles: ≤10, 10-21.6, >21.6), education, total energy intake, animal-source food consumption, sodium consumption, physical activity (tertiles: ≤57.4, 57.4-152, >152), smoking, alcohol intake, and Estimated glomerular filtration rate (eGFR). Model 2 was additionally adjusted for BMI.

"*" indicates p-value < 0.05; "**" indicates p-value<0.01

Table 5.2. Association between individual metabolites and systolic blood pressure (SBP), n=434

Metabolites	Pathway	Loading in lipid pattern*	Model 1		Model 2	
			Coefficient (95% CI)	q-value	Coefficient (95% CI)	q-value
tricosanoyl sphingomyelin (d18:1/23:0) †	Sphingomyelins	–	7.56 (4.52, 10.61)	0.002	6.53 (3.43, 9.62)	0.015
lignoceroyl sphingomyelin (d18:1/24:0) †	Sphingomyelins	–	6.37 (3.59, 9.14)	0.005	5.61 (2.83, 8.39)	0.024
palmitoylcarnitine (C16) †	Fatty Acid Metabolism (Acyl Carnitine, Long Chain Saturated)	0.50	7.26 (3.97, 10.56)	0.007	6.95 (3.71, 10.2)	0.015
1-myristoyl-2-arachidonoyl-GPC (14:0/20:4) †	Phosphatidylcholine (PC)	–	4.34 (2.32, 6.37)	0.007	3.53 (1.45, 5.61)	0.059
sphingomyelin (d18:2/24:2) †	Sphingomyelins	–	4.53 (2.41, 6.64)	0.007	4.46 (2.38, 6.54)	0.015
1-palmitoyl-2-linoleoyl-GPI (16:0/18:2) †	Phosphatidylinositol (PI)	–	5.83 (3.02, 8.64)	0.010	5.38 (2.59, 8.16)	0.037
1-palmitoyl-2-arachidonoyl-GPC (16:0/20:4n6) †	Phosphatidylcholine (PC)	–	7.91 (4.01, 11.82)	0.010	6.7 (2.77, 10.63)	0.059

1-palmitoyl-2-arachidonoyl-GPI (16:0/20:4) †	Phosphatidylinositol (PI)	–	4.96 (2.51, 7.41)	0.010	4.44 (2.01, 6.87)	0.046
sphingomyelin (d18:1/21:0, d17:1/22:0, d16:1/23:0) †	Sphingomyelins	–	6.38 (3.28, 9.49)	0.010	5.03 (1.82, 8.25)	0.088
behenoyl sphingomyelin (d18:1/22:0) †	Sphingomyelins	–	7.35 (3.57, 11.14)	0.014	5.91 (2.05, 9.77)	0.092
sphingomyelin (d18:1/14:0, d16:1/16:0) †	Sphingomyelins	–	7.21 (3.4, 11.01)	0.019	5.65 (1.75, 9.55)	0.118
cerotoylcarnitine (C26) †	Fatty Acid Metabolism (Acyl Carnitine, Long Chain Saturated)	0.42	4.29 (2.02, 6.56)	0.019	3.91 (1.66, 6.16)	0.059
pantothenate (Vitamin B5)	Pantothenate and CoA Metabolism	–	7.77 (3.63, 11.91)	0.019	6.86 (2.74, 10.97)	0.064
N2,N2-dimethylguanosine	Purine Metabolism, Guanine containing	–	8.74 (3.99, 13.49)	0.020	8.09 (3.4, 12.78)	0.059
cholesterol †	Sterol	–	8.48 (3.89, 13.07)	0.020	7.67 (3.12, 12.22)	0.062
sphingomyelin (d18:2/14:0, d18:1/14:1) †	Sphingomyelins	–	5.75 (2.58, 8.91)	0.022	4.57 (1.36, 7.79)	0.121

adrenate (22:4n6)	Long Chain Polyunsaturated Fatty Acid (n3 and n6)	0.71	3.15 (1.36, 4.93)	0.028	2.9 (1.14, 4.67)	0.066
1-palmitoleoylglycerol (16:1) †	Monoacylglycerol	0.57	2.26 (0.98, 3.53)	0.028	1.75 (0.45, 3.05)	0.143
1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1)	Phosphatidylcholine (PC)	–	3.57 (1.54, 5.59)	0.028	3.01 (0.98, 5.04)	0.107
myristoylcarnitine (C14) †	Fatty Acid Metabolism (Acyl Carnitine, Long Chain Saturated)	0.53	2.96 (1.27, 4.64)	0.029	3.01 (1.35, 4.66)	0.046
branched-chain, straight-chain, or cyclopropyl 10:1 fatty acid (1)	Partially Characterized Molecules	–	2.66 (1.13, 4.19)	0.030	2.58 (1.07, 4.08)	0.059
sphingomyelin (d18:2/16:0, d18:1/16:1) †	Sphingomyelins	–	8.35 (3.52, 13.17)	0.031	6.84 (1.99, 11.69)	0.122
picolinoylglycine	Fatty Acid Metabolism (Acyl Glycine)	–	3.34 (1.41, 5.28)	0.031	2.75 (0.81, 4.69)	0.122
N6-carbamoylthreonyladenosine	Purine Metabolism, Adenine containing	–	6.16 (2.54, 9.78)	0.032	5.57 (1.99, 9.16)	0.088
dihomolinolenate (20:3n3 or 3n6) †	Long Chain Polyunsaturated Fatty Acid (n3 and n6)	0.73	4.25 (1.75, 6.74)	0.032	3.65 (1.16, 6.14)	0.113

1-stearoyl-2-arachidonoyl-GPC (18:0/20:4)	Phosphatidylcholine (PC)		–	6.08 (2.52, 9.64)	0.032	5.01 (1.44, 8.58)	0.122
acetylcarnitine (C2)	Fatty Acid Metabolism (Acyl Carnitine, Short Chain)		0.50	6.57 (2.63, 10.5)	0.038	7 (3.13, 10.86)	0.046
retinol (Vitamin A)	Vitamin A Metabolism		–	5.16 (2.04, 8.28)	0.040	4.48 (1.39, 7.58)	0.118
argininate	Urea cycle; Arginine and Proline Metabolism		–	3.93 (1.56, 6.3)	0.040	3.19 (0.81, 5.57)	0.148
2,3-dihydroxy-5-methylthio-4-pentenoate (DMTPA)	Methionine, Cysteine, SAM and Taurine Metabolism		–	7.88 (3.11, 12.66)	0.040	6.51 (1.73, 11.29)	0.135
laurylcarnitine (C12) †	Fatty Acid Metabolism (Acyl Carnitine, Medium Chain)		0.49	2.25 (0.88, 3.62)	0.042	2.36 (1.02, 3.71)	0.059
Metabolites with 25-50% below level of detection (BLOD)/missing: Reference=BLOD/missing							
linoleoyl-linoleoyl-glycerol (18:2/18:2) [1]	Diacylglycerol	Below median	–	2.29 (-1.6, 6.17)	0.022	3.5 (-0.37, 7.37)	0.046
		Above median		-4.56 (-8.39, -0.73)		-3.45 (-7.26, 0.36)	

oleoyl-linoleoyl-glycerol (18:1/18:2) [2]	Diacylglycerol	Below median	—	-3.07 (-6.99, 0.85)	0.042	-2.49 (-6.36, 1.38)	0.088
		Above median		-7.31 (-11.39, -3.24)		-6.8 (-10.82, -2.78)	
Metabolites with >50% BLOD/missing: Reference=BLOD/missing							
phenylalanylalanine	Dipeptide	Above limit of detection	—	-4.86 (-7.86, -1.85)	0.047	-4.94 (-7.89, -1.98)	0.064

Coefficient (95% confidence interval) indicates SBP (mmHg) associated with a fold increase of the relative abundance of a given metabolite or per category change in the categorical metabolites in linear regression. The statistical significance of metabolites categorized as ordinal variables was assessed using a Wald test. Model 1 was adjusted for age, sex, provinces, batch, urbanization index (tertiles: ≤64.2, 64.2-81.5, >81.5), per capita household income (tertiles: ≤10, 10-21.6, >21.6), education, total energy intake, animal-source food consumption, sodium consumption, physical activity (tertiles: ≤57.4, 57.4-152, >152), smoking, alcohol intake, and Estimated glomerular filtration rate (eGFR). Model 2 was additionally adjusted for BMI. Metabolites are ordered by the Model 1 q-values, which are false discovery rate-adjusted p-values.

* Pattern was derived from a principal component analysis followed by a varimax rotation. Only metabolites with loading >0.4 are listed.

† Metabolites also associated with systolic blood pressure.

Table 5.3. Association between individual metabolites and diastolic blood pressure (DBP), n=434

Metabolites	**Pathway**	**Loading in lipid pattern***	**Model 1**		**Model 2**	
			Coefficient (95% CI)	**q-value**	**Coefficient (95% CI)**	**q-value**
lignoceroyl sphingomyelin (d18:1/24:0) †	Sphingomyelins	–	4.85 (3.19, 6.52)	2E-05	4.28 (2.62, 5.93)	0.001
behenoyl sphingomyelin (d18:1/22:0) †	Sphingomyelins	–	6.16 (3.9, 8.43)	1E-04	5.09 (2.8, 7.39)	0.004
tricosanoyl sphingomyelin (d18:1/23:0) †	Sphingomyelins	–	4.95 (3.11, 6.8)	1E-04	4.12 (2.26, 5.98)	0.004
cerotoylcarnitine (C26) †	Fatty Acid Metabolism (Acyl Carnitine, Long Chain Saturated)	0.42	3.09 (1.71, 4.46)	0.003	2.79 (1.45, 4.14)	0.008
corticosterone	Corticosteroids	–	-1.39 (-2.02, -0.77)	0.003	-1.11 (-1.74, -0.49)	0.033
dihomolinolenate (20:3n3 or 3n6) †	Long Chain Polyunsaturated Fatty Acid (n3 and n6)	0.73	3.28 (1.77, 4.79)	0.004	2.83 (1.34, 4.31)	0.020
sphingomyelin (d18:2/24:2) †	Sphingomyelins	–	2.8 (1.51, 4.09)	0.004	2.75 (1.5, 4)	0.004

myristoylcarnitine (C14) †	Fatty Acid Metabolism (Acyl Carnitine, Long Chain Saturated)	0.53	2.15 (1.13, 3.17)	0.005	2.19 (1.2, 3.18)	0.004
cortolone glucuronide (1)	Corticosteroids	–	2.34 (1.22, 3.45)	0.005	1.7 (0.56, 2.83)	0.112
behenoyl dihydrosphingomyelin (d18:0/22:0)	Dihydrosphingomyelins	–	1.84 (0.95, 2.72)	0.006	1.2 (0.27, 2.14)	0.182
sphingomyelin (d18:1/21:0, d17:1/22:0, d16:1/23:0) †	Sphingomyelins	–	3.85 (1.96, 5.74)	0.007	2.73 (0.79, 4.67)	0.133
1-palmitoyl-2-linoleoyl-GPI (16:0/18:2) †	Phosphatidylinositol (PI)	–	3.46 (1.75, 5.18)	0.007	3.11 (1.43, 4.79)	0.024
cis-4-decenoylcarnitine (C10:1)	Fatty Acid Metabolism (Acyl Carnitine, Monounsaturated)	0.44	1.47 (0.73, 2.21)	0.008	1.52 (0.81, 2.24)	0.006
laurylcarnitine (C12) †	Fatty Acid Metabolism (Acyl Carnitine, Medium Chain)	0.49	1.63 (0.8, 2.46)	0.009	1.72 (0.91, 2.52)	0.006
linoleate (18:2n6)	Long Chain Polyunsaturated Fatty Acid (n3 and n6)	0.80	2.65 (1.26, 4.04)	0.013	2.44 (1.08, 3.81)	0.030
decanoylcarnitine (C10)	Fatty Acid Metabolism (Acyl Carnitine, Medium Chain)	0.48	1.43 (0.67, 2.19)	0.014	1.46 (0.72, 2.2)	0.013
1-palmitoleoylglycerol (16:1) †	Monoacylglycerol	0.57	1.45 (0.68, 2.23)	0.014	1.05 (0.27, 1.84)	0.157

palmitate (16:0)	Long Chain Saturated Fatty Acid	0.80	3.51 (1.62, 5.41)	0.014	3.11 (1.26, 4.97)	0.054
cis-4-decenoate	Medium Chain Fatty Acid	—	2.06 (0.95, 3.17)	0.014	2.03 (0.95, 3.11)	0.021
1-myristoyl-2-arachidonoyl-GPC (14:0/20:4) †	Phosphatidylcholine (PC)	—	2.29 (1.05, 3.53)	0.014	1.59 (0.33, 2.85)	0.188
5-dodecenoylcarnitine (C12:1)	Fatty Acid Metabolism (Acyl Carnitine, Monounsaturated)	0.50	1.43 (0.66, 2.2)	0.014	1.53 (0.78, 2.29)	0.009
1-dihomo-linoleoylglycerol (20:2)	Monoacylglycerol	0.64	1.39 (0.62, 2.15)	0.018	1.06 (0.3, 1.83)	0.141
octanoylcarnitine (C8)	Fatty Acid Metabolism (Acyl Carnitine, Medium Chain)	0.51	1.74 (0.78, 2.71)	0.018	1.83 (0.9, 2.77)	0.014
1-linoleoylglycerol (18:2)	Monoacylglycerol	0.63	1.54 (0.68, 2.39)	0.018	1.14 (0.28, 2)	0.161
1-dihomo-linolenylglycerol (20:3)	Monoacylglycerol	0.66	1.49 (0.65, 2.33)	0.020	1.02 (0.17, 1.87)	0.225
palmitoylcarnitine (C16) †	Fatty Acid Metabolism (Acyl Carnitine, Long Chain Saturated)	0.50	3.57 (1.55, 5.59)	0.021	3.33 (1.36, 5.3)	0.051
tetrahydrocortisone glucuronide (5)	Corticosteroids	—	1.65 (0.7, 2.6)	0.024	1.22 (0.27, 2.17)	0.182

1-palmitoyl-2-arachidonoyl-GPI (16:0/20:4) †	Phosphatidylinositol (PI)	–	2.6 (1.11, 4.1)	0.024	2.19 (0.72, 3.67)	0.112
sphingomyelin (d18:2/16:0, d18:1/16:1) †	Sphingomyelins	–	5.09 (2.15, 8.03)	0.024	3.9 (0.97, 6.82)	0.161
hexanoylcarnitine (C6)	Fatty Acid Metabolism (Acyl Carnitine, Medium Chain)	0.55	1.79 (0.75, 2.83)	0.025	1.83 (0.82, 2.84)	0.027
sphingomyelin (d18:2/14:0, d18:1/14:1) †	Sphingomyelins	–	3.31 (1.38, 5.24)	0.025	2.36 (0.42, 4.3)	0.217
1-arachidonoyl-GPI (20:4)	Lysophospholipid	0.55	3.53 (1.46, 5.6)	0.026	2.99 (0.95, 5.02)	0.118
sphingomyelin (d18:1/14:0, d16:1/16:0) †	Sphingomyelins	–	3.95 (1.63, 6.27)	0.026	2.65 (0.29, 5.01)	0.279
cholesterol †	Sterol	–	4.78 (1.98, 7.59)	0.026	4.15 (1.4, 6.9)	0.112
2-palmitoleoylglycerol (16:1)	Monoacylglycerol	0.46	1.14 (0.46, 1.82)	0.031	0.81 (0.13, 1.5)	0.237
hydantoin-5-propionate	Histidine Metabolism	–	1.45 (0.58, 2.32)	0.032	1.32 (0.47, 2.17)	0.094
1-palmitoyl-2-arachidonoyl-GPC (16:0/20:4n6) †	Phosphatidylcholine (PC)	–	3.94 (1.55, 6.33)	0.035	2.94 (0.55, 5.32)	0.206
p-cresol sulfate	Benzoate Metabolism	–	-1.01 (-1.63, -0.4)	0.036	-0.88 (-1.48, -0.27)	0.118

palmitoleate (16:1n7)	Long Chain Monounsaturated Fatty Acid	0.70	1.8 (0.69, 2.91)	0.040	1.71 (0.63, 2.79)	0.077

Coefficient (95% confidence interval) indicates DBP (mmHg) associated with a fold increase of the relative abundance of a given metabolite in linear regression. Model 1 was adjusted for age, sex, provinces, batch, urbanization index (tertiles: ≤64.2, 64.2-81.5, >81.5), per capita household income (tertiles: ≤10, 10-21.6, >21.6), education, total energy intake, animal-source food consumption, sodium consumption, physical activity (tertiles: ≤57.4, 57.4-152, >152), smoking, alcohol intake, and Estimated glomerular filtration rate (eGFR). Model 2 was additionally adjusted for BMI. Metabolites are ordered by the Model 1 q-values, which are false discovery rate-adjusted p-values.

[*] Pattern was derived from principal component analysis followed by a varix rotation. Only metabolites with loading > 0.4 are listed.

[†] Metabolites also associated with diastolic blood pressure.

Table 5.4. Metabolic pathway analysis

		Systolic blood pressure (SBP)				Diastolic blood pressure (DBP)			
	m*	**k***	**Enrichment score***	**p-value†**	**q-value†**	**k***	**Enrichment score***	**p-value†**	**q-value†**
Benzoate Metabolism	30	0	--	--	--	1	0.77	1	1
Corticosteroids	6	0	--	--	--	3	12.47	0.001	0.005
Diacylglycerol	3	2	16.23	0.004	0.019	0	--	--	--
Dihydrosphingomyelins	5	0	--	--	--	1	4.73	0.198	0.337
Dipeptide	14	1	1.67	0.418	0.496	0	--	--	--
Fatty Acid Metabolism (Acyl Carnitine, Long Chain Saturated)	6	3	14.48	9E-04	0.009	3	14.48	0.001	0.005
Fatty Acid Metabolism (Acyl Carnitine, Medium Chain)	6	1	4.54	0.206	0.301	4	19.96	4E-05	3E-04
Fatty Acid Metabolism (Acyl Carnitine, Monounsaturated)	9	0	--	--	--	2	6.22	0.054	0.131
Fatty Acid Metabolism (Acyl Carnitine, Short Chain)	1	1	27.36	0.038	0.103	0	--	--	--
Fatty Acid Metabolism (Acyl Glycine)	5	1	4.73	0.175	0.277	0	--	--	--

Histidine Metabolism	16	0	--	--	--	1	1.46	0.509	0.577
Long Chain Monounsaturated Fatty Acid	7	0	--	--	--	1	3.37	0.266	0.348
Long Chain Polyunsaturated Fatty Acid (n3 and n6)	16	2	3.00	0.119	0.270	2	3.00	0.149	0.317
Long Chain Saturated Fatty Acid	7	0	--	--	--	1	3.37	0.266	0.348
Lysophospholipid	27	0	--	--	--	1	0.85	1	1
Medium Chain Fatty Acid	11	0	--	--	--	1	2.14	0.386	0.469
Methionine, Cysteine, SAM and Taurine Metabolism	22	1	1.06	0.574	0.574	0	--	--	--
Monoacylglycerol	14	1	1.67	0.418	0.496	5	9.35	0.0002	0.001
Pantothenate and CoA Metabolism	1	1	23.76	0.038	0.103	0	--	--	--
Partially Characterized Molecule	17	1	1.37	0.482	0.539	0	--	--	--
Phosphatidylcholine (PC)	18	4	5.63	0.004	0.019	2	2.66	0.180	0.337
Phosphatidylinositol (PI)	5	2	9.72	0.013	0.049	2	9.72	0.017	0.048
Purine Metabolism, Adenine containing	5	1	4.73	0.175	0.277	0	--	--	--
Purine Metabolism, Guanine containing	4	1	5.92	0.142	0.270	0	--	--	--
Sphingomyelins	28	8	8.07	3E-06	6E-05	8	8.07	1E-05	2E-04
Sterol	7	1	3.37	0.236	0.320	1	3.37	0.266	0.348

Urea cycle; Arginine and Proline Metabolism	21	1	1.11	0.557	0.574	0	--	--
Vitamin A Metabolism	4	1	5.92	0.142	0.270	0	--	--

[*] Enrichment score was calculated using (k/m)/[(n-k)/(N-m)], where k is the number of metabolites associated with SBP or DBP in the pathway (false discovery rate adjusted p-value in linear regression Model 1, q-value<0.05), m is the total number of tested metabolites in the pathway, n is the total number of known metabolites associated with SBP (n=34) or DBP (n=39), and N is the total number of known metabolites (N=904).

[†] P-value for each pathway was calculated using Fisher's exact test and false discovery rate adjusted (q-value) across eligible pathways (i.e., containing at least one metabolite associated with SBP or DBP).

Figure 5.1. Microbial between-person diversity (β-diversity) assessed using principal coordinate analysis (PCoA) by systolic blood pressure (SBP). MDS, multidivisional scaling (i.e., PCoA axis). For better visualization of the separation, SBP (mmHg) was categorized to quartiles. Plot legend shows the color and range of SBP for each quartile. Centroids illustrate the 95% CI for the mean location of each SBP quartile. Because SBP show separation along MDS4 only, the 10 taxa had the greatest contribution to MDS4 are shown as vectors, indicating directions and strengths (vector lengths) of their associations along PCoA axes. Table S5.1 showed the full list of taxa scores for each axis. In permutational multivariate analysis of variance (PERMANOVA), SBP (continuous) had R^2 of 2.01% and p-value of 0.002, after adjusting for age, sex, provinces, urbanization index (tertiles: ≤64.2, 64.2-81.5, >81.5), per capita household income (tertiles: ≤10, 10-21.6, >21.6), education, total energy intake, animal-source food consumption, sodium consumption, physical activity (tertiles: ≤57.4, 57.4-152, >152), smoking, alcohol intake, and estimated glomerular filtration rate. Results remained the same after additionally adjustment of BMI.

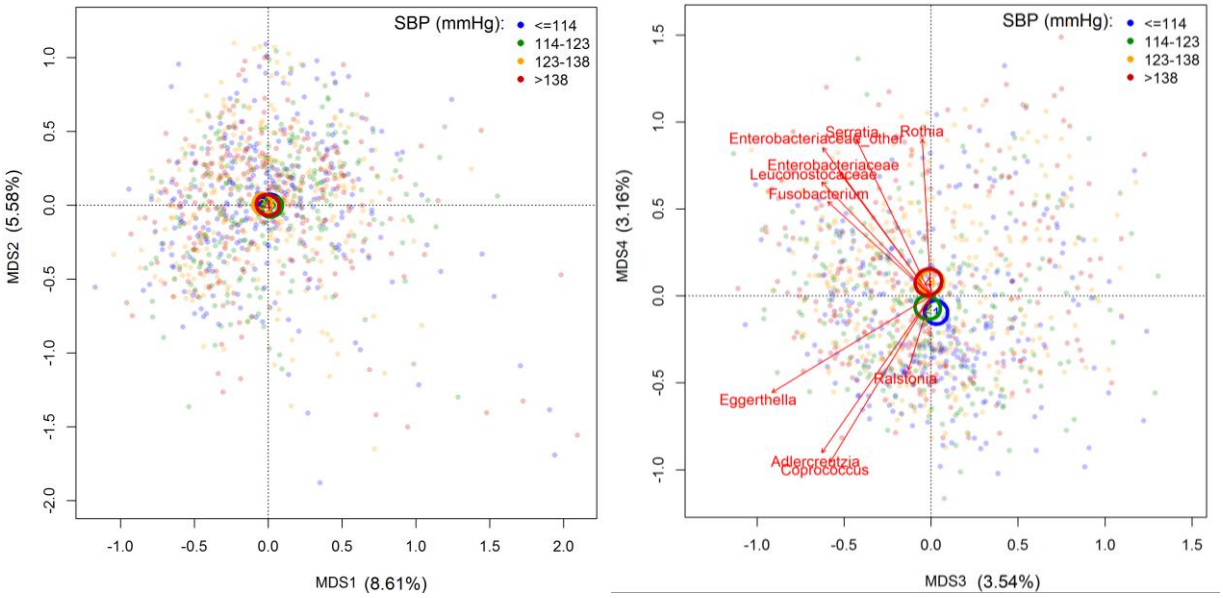


Table S5.1. Taxa scores for principal coordinate analysis (PCoA) axes (n=1,003)

MDS1	MDS2	MDS3	MDS4	phylum	class	order	family	genus
0.44	0.45	-	-	Euryarchaeota	Methanobacteria	Methanobacteriales	Methanobacteriacea e	Methanobrevibacte r
-	-	-	0.67	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	Rothia
0.48	-	-	-	Actinobacteria	Actinobacteria	Actinomycetales	Nocardiaceae	Rhodococcus
-0.70	-	-0.64	-	Actinobacteria	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium
-	-	0.57	-0.86	Actinobacteria	Coriobacteriia	Coriobacteriales	Coriobacteriaceae	Adlercreutzia
-	-	0.80	-0.51	Actinobacteria	Coriobacteriia	Coriobacteriales	Coriobacteriaceae	Eggerthella
0.45	-	-	-	Bacteroidetes	Bacteroidia	Bacteroidales		
0.46	-	-	-	Bacteroidetes	Bacteroidia	Bacteroidales	[Barnesiellaceae]	
0.42	-	-	-	Bacteroidetes	Bacteroidia	Bacteroidales	[Odoribacteraceae]	Odoribacter
0.74	-	-	-	Bacteroidetes	Bacteroidia	Bacteroidales	[Paraprevotellaceae]	[Prevotella]
0.96	0.85	-	-	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
1.01	0.67	-	-	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadacea e	Parabacteroides
1.33	0.70	-	-	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella
0.77	-	-	-	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	

1.20	-	-	-	Bacteroidetes	Bacteroidia	Bacteroidales	S24-7	
-	-	-0.67	-	Cyanobacteria	4C0d-2	YS2		
0.45	-	-	-	Firmicutes	Bacilli	Bacillales		
-	-	-	0.49	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus
-	-0.63	0.73	0.80	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	
-	-0.69	0.93	-	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Lactococcus
0.45	0.50	-	-	Firmicutes	Clostridia	Clostridiales	Christensenellaceae	
-	-	0.49	-	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium
-	-	0.70	-	Firmicutes	Clostridia	Clostridiales	Eubacteriaceae	Pseudoramibacter_ Eubacterium
-	0.51	-	-	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Anaerostipes
-	-	0.43	-0.41	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Blautia
-	-	-	-0.52	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Coprococcus
-	0.83	-	-	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Lachnospira
-	0.46	-	-	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Roseburia
-	-	0.63	-	Firmicutes	Clostridia	Clostridiales	Peptococcaceae	Peptococcus
-	-0.43	-	-	Firmicutes	Clostridia	Clostridiales	Peptostreptococcaceae	
-	0.55	-	-	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	

0.51	0.80	-	-	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Faecalibacterium
0.57	0.50	-	-	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Oscillospira
-	0.45	-	-	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Ruminococcus
0.52	0.67	-	-	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Phascolarctobacterium
-	-	-	0.48	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Veillonella
0.45	-	-	-	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	[Eubacterium]
0.58	-	-0.74	-	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	Catenibacterium
-	-	0.44	-	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	Coprobacillus
-	0.49	-	-	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	Holdemania
-	-	0.57	0.54	Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Fusobacterium
0.56	-	-	-	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	
0.54	-	-	-	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	
0.62	-	-	-	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	Sutterella

0.80	-0.70	-	-	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	
1.06	-1.01	-	-	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Delftia
0.40	-0.46	-	-	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Other
1.24	-1.03	-	-0.50	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Ralstonia
0.43	-	-	-	Proteobacteria	Deltaproteobacteria	Desulfovibrionales	Desulfovibrionaceae	Desulfovibrio
-	-	0.50	0.71	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	
-	-	-	0.71	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Serratia
-	-	0.56	0.81	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Other
0.65	-0.48	-	-	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter

0.51	-0.63	-	-	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	
0.74	0.42	-	-0.41	Tenericutes	Mollicutes	RF39		
0.44	-	-	-	TM7	TM7-3			
-	0.79	-	-	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiacea e	Akkermansia
0.84	-	-	-	Other	Other	Other	Other	Other

MDS, multidimensional scaling. PCoA was performed for 1,008 taxa based on Bray-Curtis distance matrix. The first four axes each explained 8.61, 5.58, 3.54, and 3.12 percent variance of gut microbiota, respectively. Taxa with scores ≥ 0.40 or ≤ -0.40 are shown here. Taxa are ordered alphabetically.

Table S5.2. Metabolites loadings for principal component analysis (PCA) axes (n=434)

Axis 1	Axis 2	Axis 3	Metabolites	Class	Pathway
			(2 or 3)-decenoate (10:1n8 or n7)	Lipid	Medium Chain Fatty Acid
-	0.58	-	(2,4 or 2,5)-dimethylphenol sulfate	Xenobiotics	Food Component/Plant
0.45	-	-	(R)-3-hydroxybutyrylcarnitine	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Hydroxy)
-	0.60	-	(S)-3-hydroxybutyrylcarnitine	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Hydroxy)
-	0.45	-	(S)-a-Amino-omega-caprolactam	Xenobiotics	Food Component/Plant
0.44	-	-	1-(1-enyl-oleoyl)-GPE (P-18:1)*	Lipid	Lysoplasmalogen
-	-	0.59	1-(1-enyl-palmitoyl)-GPC (P-16:0)*	Lipid	Lysoplasmalogen
-	-	0.52	1-(1-enyl-palmitoyl)-GPE (P-16:0)*	Lipid	Lysoplasmalogen
-	-	0.59	1-(1-enyl-stearoyl)-GPE (P-18:0)*	Lipid	Lysoplasmalogen
-	-	0.54	10-heptadecenoate (17:1n7)	Lipid	Long Chain Monounsaturated Fatty Acid
-	0.79	-	10-nonadecenoate (19:1n9)	Lipid	Long Chain Monounsaturated Fatty Acid
-	0.74	-	10-undecenoate (11:1n1)	Lipid	Medium Chain Fatty Acid
-	0.47	-	13-HODE + 9-HODE	Lipid	Fatty Acid, Monohydroxy
-	-	0.51	16-hydroxypalmitate	Lipid	Fatty Acid, Monohydroxy
-	0.56	-	17alpha-hydroxypregnenolone 3-sulfate	Lipid	Pregnenolone Steroids
-	-	-0.41	1-arachidonoyl-GPI* (20:4)*	Lipid	Lysophospholipid

-	0.55	-	1-arachidonylglycerol (20:4)	Lipid	Monoacylglycerol
-	0.62	-	1-carboxyethylisoleucine	Amino Acid	Leucine, Isoleucine and Valine Metabolism
0.60	-	-	1-carboxyethylleucine	Amino Acid	Leucine, Isoleucine and Valine Metabolism
0.54	-	-	1-carboxyethylphenylalanine	Amino Acid	Phenylalanine Metabolism
0.66	-	-	1-carboxyethyltyrosine	Amino Acid	Tyrosine Metabolism
0.54	-	-	1-carboxyethylvaline	Amino Acid	Leucine, Isoleucine and Valine Metabolism
0.55	-	-	1-dihomo-linolenylglycerol (20:3)	Lipid	Monoacylglycerol
-	0.66	-	1-dihomo-linoleoylglycerol (20:2)	Lipid	Monoacylglycerol
-	0.64	-	1-docosahexaenoylglycerol (22:6)	Lipid	Monoacylglycerol
-	0.49	-	1-linolenoylglycerol (18:3)	Lipid	Monoacylglycerol
-	0.56	-	1-linoleoylglycerol (18:2)	Lipid	Monoacylglycerol
-	0.63	-	1-linoleoyl-GPE (18:2)*	Lipid	Lysophospholipid
0.52	-	-	1-methyl-4-imidazoleacetate	Amino Acid	Histidine Metabolism
0.55	-	-	1-methyladenosine	Nucleotide	Purine Metabolism, Adenine containing
0.44	-	-	1-methylhistidine	Amino Acid	Histidine Metabolism
0.62	-	-	1-myristoylglycerol (14:0)	Lipid	Monoacylglycerol
-	0.43	-	1-oleoylglycerol (18:1)	Lipid	Monoacylglycerol
-	0.67	-	1-oleoyl-GPC (18:1)	Lipid	Lysophospholipid

0.42	-	-	1-oleoyl-GPE (18:1)	Lipid	Lysophospholipid
0.61	-	-	1-palmitoleoylglycerol (16:1)*	Lipid	Monoacylglycerol
-	0.56	-	1-palmitoyl-2-linoleoyl-GPE (16:0/18:2)	Lipid	Phosphatidylethanolamine (PE)
0.42	-	-	1-palmitoyl-2-oleoyl-GPE (16:0/18:1)	Lipid	Phosphatidylethanolamine (PE)
0.43	-	-	1-palmitoylglycerol (16:0)	Lipid	Monoacylglycerol
-	0.57	-	1-palmitoyl-GPI* (16:0)	Lipid	Lysophospholipid
-	-	0.46	1-ribosyl-imidazoleacetate*	Amino Acid	Histidine Metabolism
0.44	-	-	1-stearoyl-2-linoleoyl-GPE (18:0/18:2)*	Lipid	Phosphatidylethanolamine (PE)
0.43	-	-	1-stearoyl-2-oleoyl-GPC (18:0/18:1)	Lipid	Phosphatidylcholine (PC)
0.45	-	-	1-stearoyl-2-oleoyl-GPE (18:0/18:1)	Lipid	Phosphatidylethanolamine (PE)
0.48	-	-	1-stearoyl-GPC (18:0)	Lipid	Lysophospholipid
-	-	0.51	1-stearoyl-GPI (18:0)	Lipid	Lysophospholipid
-	-	0.52	2,3-dihydroxy-5-methylthio-4-pentenoate (DMTPA)*	Amino Acid	Methionine, Cysteine, SAM and Taurine Metabolism
0.76	-	-	2-arachidonoylglycerol (20:4)	Lipid	Monoacylglycerol
-	0.40	-	2-hydroxy-3-methylvalerate	Amino Acid	Leucine, Isoleucine and Valine Metabolism
0.48	-	-	2-hydroxybutyrate/2-hydroxyisobutyrate	Amino Acid	Glutathione Metabolism
-	0.57	-	2-hydroxyfluorene sulfate	Xenobiotics	Tobacco Metabolite

0.45	-	-	2-hydroxypalmitate	Lipid	Fatty Acid, Monohydroxy
-	0.48	-	2-hydroxysebacic acid	Lipid	Fatty Acid, Dicarboxylate
0.45	-	-	2-ketocaprylate	Amino Acid	Leucine, Isoleucine and Valine Metabolism
-	-	-0.55	2-linoleoylglycerol (18:2)	Lipid	Monoacylglycerol
-	0.58	-	2-methylbutyrylcarnitine (C5)	Amino Acid	Leucine, Isoleucine and Valine Metabolism
0.55	-	-	2-oleoylglycerol (18:1)	Lipid	Monoacylglycerol
-	0.65	-	2-palmitoleoylglycerol (16:1)*	Lipid	Monoacylglycerol
-	0.46	-	3-(3-amino-3-carboxypropyl)uridine*	Nucleotide	Pyrimidine Metabolism, Uracil containing
0.69	-	-	3-(4-hydroxyphenyl)lactate (HPLA)	Amino Acid	Tyrosine Metabolism
0.58	-	-	3-amino-2-piperidone	Amino Acid	Urea cycle; Arginine and Proline Metabolism
0.52	-	-	3-hydroxy-2-ethylpropionate	Amino Acid	Leucine, Isoleucine and Valine Metabolism
0.54	-	-	3-hydroxy-3-methylglutarate	Lipid	Mevalonate Metabolism
0.67	-	-	3-hydroxybutyrate (BHBA)	Lipid	Ketone Bodies
-	0.60	-	3-hydroxydecanoate	Lipid	Fatty Acid, Monohydroxy
-	0.66	-	3-hydroxyhexanoate	Lipid	Fatty Acid, Monohydroxy
-	0.56	-	3-hydroxylaurate	Lipid	Fatty Acid, Monohydroxy
-	0.63	-	3-hydroxyoctanoate	Lipid	Fatty Acid, Monohydroxy
-	0.56	-	3-hydroxyoctanoylcarnitine (1)	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Hydroxy)

-	0.56	-	3-hydroxyoctanoylcarnitine (2)	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Hydroxy)
-	0.56	-	3-hydroxypyridine sulfate	Xenobiotics	Chemical
0.53	-	-	3-indoleglyoxylic acid	Xenobiotics	Food Component/Plant
0.41	-	-	3-methoxytyramine sulfate	Amino Acid	Tyrosine Metabolism
0.57	-	-	3-methyl catechol sulfate (1)	Xenobiotics	Benzoate Metabolism
0.47	-	-	3-methyl-2-oxobutyrate	Amino Acid	Leucine, Isoleucine and Valine Metabolism
-	-	-0.65	3-methyl-2-oxovalerate	Amino Acid	Leucine, Isoleucine and Valine Metabolism
-	-	-0.59	4-acetylphenyl sulfate	Xenobiotics	Benzoate Metabolism
0.43	-	-	4-guanidinobutanoate	Amino Acid	Guanidino and Acetamido Metabolism
-	-	0.51	4-HDoHE	Lipid	Docosanoid
-	-	0.59	4-methyl-2-oxopentanoate	Amino Acid	Leucine, Isoleucine and Valine Metabolism
-	-	-0.60	4-vinylphenol sulfate	Xenobiotics	Benzoate Metabolism
0.53	-	-	5-(galactosylhydroxy)-L-lysine	Amino Acid	Lysine Metabolism
0.65	-	-	5,6-dihydrouridine	Nucleotide	Pyrimidine Metabolism, Uracil containing
0.70	-	-	5alpha-androstan-3beta,17beta-diol disulfate	Lipid	Androgenic Steroids
-	-	-0.45	5-dodecenoate (12:1n7)	Lipid	Medium Chain Fatty Acid

-	0.58	-	5-dodecenoylcarnitine (C12:1)	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Monounsaturated)
-	0.50	-	5-HEPE	Lipid	Eicosanoid
-	-	0.57	5-HETE	Lipid	Eicosanoid
-	-	0.61	5-HETrE	Lipid	Eicosanoid
-	-	0.59	5-hydroxylysine	Amino Acid	Lysine Metabolism
0.41	-	-	5-methylthioribose	Amino Acid	Methionine, Cysteine, SAM and Taurine Metabolism
0.44	-	-	5-methyluridine (ribothymidine)	Nucleotide	Pyrimidine Metabolism, Uracil containing
0.48	-	-	7-methylguanine	Nucleotide	Purine Metabolism, Guanine containing
0.53	-	-	acetylcarnitine (C2)	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Short Chain)
-	0.50	-	acisoga	Amino Acid	Polyamine Metabolism
0.40	-	-	adrenate (22:4n6)	Lipid	Long Chain Polyunsaturated Fatty Acid (n3 and n6)
-	0.71	-	alpha-hydroxyisovalerate	Amino Acid	Leucine, Isoleucine and Valine Metabolism
0.44	-	-	alpha-ketoglutarate	Energy	TCA Cycle

-	-	-0.57	alpha-tocopherol	Cofactors and Vitamins	Tocopherol Metabolism
-	-	-0.48	androstenediol (3alpha, 17alpha) monosulfate (3)	Lipid	Androgenic Steroids
-	-	-0.49	androstenediol (3beta,17beta) disulfate (1)	Lipid	Androgenic Steroids
-	-	-0.50	androstenediol (3beta,17beta) disulfate (2)	Lipid	Androgenic Steroids
-	0.41	-0.47	androstenediol (3beta,17beta) monosulfate (1)	Lipid	Androgenic Steroids
-	-	-0.48	arabitol/xylitol	Carbohydrate	Pentose Metabolism
0.57	-	-	arachidate (20:0)	Lipid	Long Chain Saturated Fatty Acid
-	0.47	-	arachidonate (20:4n6)	Lipid	Long Chain Polyunsaturated Fatty Acid (n3 and n6)
-	0.66	-	argininate*	Amino Acid	Urea cycle; Arginine and Proline Metabolism
0.46	-	-	ascorbic acid 3-sulfate*	Cofactors and Vitamins	Ascorbate and Aldarate Metabolism
0.49	-	-	behenoyl sphingomyelin (d18:1/22:0)*	Lipid	Sphingomyelins
-	-	0.40	beta-citrylglutamate	Amino Acid	Glutamate Metabolism

0.43	-	-	bilirubin	Cofactors and Vitamins	Hemoglobin and Porphyrin Metabolism
-	-	-0.52	bilirubin (E,E)*	Cofactors and Vitamins	Hemoglobin and Porphyrin Metabolism
-	-	-0.45	bilirubin degradation product, C17H18N2O4 (2)**	Partially Characterized Molecules	Partially Characterized Molecules
-	-	-0.49	branched-chain, straight-chain, or cyclopropyl 12:1 fatty acid*	Partially Characterized Molecules	Partially Characterized Molecules
-	0.62	-	butyrate/isobutyrate (4:0)	Lipid	Short Chain Fatty Acid
-	-	0.60	caprate (10:0)	Lipid	Medium Chain Fatty Acid
-	0.49	-	caproate (6:0)	Lipid	Medium Chain Fatty Acid
-	-	0.46	cerotoylcarnitine (C26)*	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Long Chain Saturated)
-	0.42	-	C-glycosyltryptophan	Amino Acid	Tryptophan Metabolism
0.72	-	-	cholesterol	Lipid	Sterol
-	-	0.42	choline	Lipid	Phospholipid Metabolism

0.47	-	-	cis-4-decenoate	Lipid	Medium Chain Fatty Acid
-	0.61	-	cis-4-decenoylcarnitine (C10:1)	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Monounsaturated)
-	0.44	-	citrulline	Amino Acid	Urea cycle; Arginine and Proline Metabolism
0.54	-	-	creatinine	Amino Acid	Creatine Metabolism
0.61	-	-	cystathionine	Amino Acid	Methionine, Cysteine, SAM and Taurine Metabolism
0.47	-	-	cysteine	Amino Acid	Methionine, Cysteine, SAM and Taurine Metabolism
-	-	-0.53	cysteine s-sulfate	Amino Acid	Methionine, Cysteine, SAM and Taurine Metabolism
-		0.47	decanoylcarnitine (C10)	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Medium Chain)
-	0.48	-	dehydroepiandrosterone sulfate (DHEA-S)	Lipid	Androgenic Steroids
-	-	-0.42	dihomolinoleate (20:2n6)	Lipid	Long Chain Polyunsaturated Fatty Acid (n3 and n6)

-	0.74	-	dihomolinolenate (20:3n3 or 3n6)	Lipid	Long Chain Polyunsaturated Fatty Acid (n3 and n6)
-	0.73	-	dimethylarginine (ADMA + SDMA)	Amino Acid	Urea cycle; Arginine and Proline Metabolism
0.70	-	-	docosadienoate (22:2n6)	Lipid	Long Chain Polyunsaturated Fatty Acid (n3 and n6)
-	0.50	-	docosahexaenoate (DHA; 22:6n3)	Lipid	Long Chain Polyunsaturated Fatty Acid (n3 and n6)
-	0.70	-	docosapentaenoate (DPA; 22:5n3)	Lipid	Long Chain Polyunsaturated Fatty Acid (n3 and n6)
-	0.78	-	docosatrienoate (22:3n3)	Lipid	Long Chain Polyunsaturated Fatty Acid (n3 and n6)
-	0.54	-	dodecadienoate (12:2)*	Lipid	Fatty Acid, Dicarboxylate
-	0.62	-	dopamine 3-O-sulfate	Amino Acid	Tyrosine Metabolism
0.47	-	-	eicosapentaenoate (EPA; 20:5n3)	Lipid	Long Chain Polyunsaturated Fatty Acid (n3 and n6)
-	0.47	-	eicosenoate (20:1n9 or 1n11)	Lipid	Long Chain Monounsaturated Fatty Acid
-	0.57	-	formiminoglutamate	Amino Acid	Histidine Metabolism
0.52	-	-	glucuronate	Carbohydrate	Aminosugar Metabolism

0.42	-	-	glutamate	Amino Acid	Glutamate Metabolism
-	0.42	-	glutamine conjugate of C6H10O2 (2)*	Partially Characterized Molecules	Partially Characterized Molecules
-	0.47	-	glutaroylcarnitine (C5)	Amino Acid	Lysine Metabolism
0.41	-	-	glycerol	Lipid	Glycerolipid Metabolism
-	0.65	-	glycerophosphoethanolamine	Lipid	Phospholipid Metabolism
-	-	0.53	guaiacol sulfate	Xenobiotics	Benzoate Metabolism
0.43	-	-	heptanoate (7:0)	Lipid	Medium Chain Fatty Acid
-	-	0.42	hexadecadienoate (16:2n6)	Lipid	Long Chain Polyunsaturated Fatty Acid (n3 and n6)
-	0.74	-	hexadecanedioate (C16)	Lipid	Fatty Acid, Dicarboxylate
-	0.47	-	hexanoylcarnitine (C6)	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Medium Chain)
-	0.54	-	hexanoylglutamine	Lipid	Fatty Acid Metabolism (Acyl Glutamine)
-	0.52	-	homocitrulline	Amino Acid	Urea cycle; Arginine and Proline Metabolism
0.57	-	-	hydroxyasparagine	Amino Acid	Alanine and Aspartate Metabolism
0.69	-	-	hydroxy-N6,N6,N6-trimethyllysine*	Amino Acid	Lysine Metabolism

0.71	-	-	hydroxypalmitoyl sphingomyelin (d18:1/16:0(OH))**	Lipid	Sphingomyelins
-	-	0.42	imidazole lactate	Amino Acid	Histidine Metabolism
0.46	-	-	indolelactate	Amino Acid	Tryptophan Metabolism
0.64	-	-	isoleucine	Amino Acid	Leucine, Isoleucine and Valine Metabolism
0.42	-	-	isovalerate (C5)	Amino Acid	Leucine, Isoleucine and Valine Metabolism
-	-	0.59	kynurenine	Amino Acid	Tryptophan Metabolism
0.59	-	-	lactate	Carbohydrate	Glycolysis, Gluconeogenesis, and Pyruvate Metabolism
0.45	-	-	laurate (12:0)	Lipid	Medium Chain Fatty Acid
-	0.43	-	laurylcarnitine (C12)	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Medium Chain)
-	0.49	-	leukotriene B4	Lipid	Eicosanoid
-	-	0.63	linoleate (18:2n6)	Lipid	Long Chain Polyunsaturated Fatty Acid (n3 and n6)
-	0.80	-	linolenate (18:3n3 or 3n6)	Lipid	Long Chain Polyunsaturated Fatty Acid (n3 and n6)
-	0.72	-	linoleoyl ethanolamide	Lipid	Endocannabinoid

-	0.54	-	linoleoylcarnitine (C18:2)*	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Polyunsaturated)
-	0.41	-	margarate (17:0)	Lipid	Long Chain Saturated Fatty Acid
-	0.74	-	methionine sulfone	Amino Acid	Methionine, Cysteine, SAM and Taurine Metabolism
0.44	-	-	methionine sulfoxide	Amino Acid	Methionine, Cysteine, SAM and Taurine Metabolism
-	-	0.61	methylnaphthyl sulfate (2)*	Xenobiotics	Chemical
0.44	-	-	myristate (14:0)	Lipid	Long Chain Saturated Fatty Acid
-	0.70	-	myristoleate (14:1n5)	Lipid	Long Chain Monounsaturated Fatty Acid
-	0.61	-	myristoleoylcarnitine (C14:1)*	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Monounsaturated)
-	0.51	-	myristoyl dihydrosphingomyelin (d18:0/14:0)*	Lipid	Dihydrosphingomyelins
-	-	0.47	myristoylcarnitine (C14)	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Long Chain Saturated)
-	0.53	-	N,N,N-trimethyl-alanylproline betaine (TMAP)	Amino Acid	Urea cycle; Arginine and Proline Metabolism

0.60	-	-	N,N-dimethyl-pro-pro	Peptide	Modified Peptides
0.73	-	-	N1-methylinosine	Nucleotide	Purine Metabolism, (Hypo)Xanthine/Inosine containing
0.57	-	-	N2,N2-dimethylguanosine	Nucleotide	Purine Metabolism, Guanine containing
0.60	-	-	N2,N5-diacetylornithine	Amino Acid	Urea cycle; Arginine and Proline Metabolism
0.56	-	-	N6,N6,N6-trimethyllysine	Amino Acid	Lysine Metabolism
0.58	-	-	N6-acetyllysine	Amino Acid	Lysine Metabolism
0.69	-	-	N6-carbamoylthreonyladenosine	Nucleotide	Purine Metabolism, Adenine containing
0.65	-	-	N-acetylalanine	Amino Acid	Alanine and Aspartate Metabolism
0.69	-	-	N-acetylglucosamine/N-acetylgalactosamine	Carbohydrate	Aminosugar Metabolism
0.65	-	-	N-acetyl-isoputreanine*	Amino Acid	Polyamine Metabolism
0.42	-	-	N-acetylneuraminate	Carbohydrate	Aminosugar Metabolism
0.65	-	-	N-acetylputrescine	Amino Acid	Polyamine Metabolism
0.50	-	-	N-acetylserine	Amino Acid	Glycine, Serine and Threonine Metabolism
0.67	-	-	N-acetyltaurine	Amino Acid	Methionine, Cysteine, SAM and Taurine Metabolism
0.66	-	-	N-acetylthreonine	Amino Acid	Glycine, Serine and Threonine Metabolism

0.73	-	-	N-acetylvaline	Amino Acid	Leucine, Isoleucine and Valine Metabolism
0.42	-	-	N-delta-acetylornithine	Amino Acid	Urea cycle; Arginine and Proline Metabolism
0.50	-	-	N-formylmethionine	Amino Acid	Methionine, Cysteine, SAM and Taurine Metabolism
0.55	-	-	N-linoleoyltaurine*	Lipid	Endocannabinoid
-	0.46	-	N-oleoyltaurine	Lipid	Endocannabinoid
-	0.63	-	nonadecanoate (19:0)	Lipid	Long Chain Saturated Fatty Acid
-	0.58	-	o-cresol sulfate	Xenobiotics	Benzoate Metabolism
0.50	-	-	octadecanedioate (C18)	Lipid	Fatty Acid, Dicarboxylate
-	0.48	-	octanoylcarnitine (C8)	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Medium Chain)
-	0.51	-	oleate/vaccenate (18:1)	Lipid	Long Chain Monounsaturated Fatty Acid
-	0.80	-	oleoyl ethanolamide	Lipid	Endocannabinoid
-	0.66	-	oleoylcarnitine (C18)	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Monounsaturated)
-	0.54	-	ornithine	Amino Acid	Urea cycle; Arginine and Proline Metabolism
0.51	-	-	orotidine	Nucleotide	Pyrimidine Metabolism, Orotate containing
0.71	-	-	O-sulfo-L-tyrosine	Xenobiotics	Chemical

0.63	-	-	palmitate (16:0)	Lipid	Long Chain Saturated Fatty Acid
-	0.80	-	palmitoleate (16:1n7)	Lipid	Long Chain Monounsaturated Fatty Acid
-	0.70	-	palmitoleoylcarnitine (C16:1)*	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Monounsaturated)
-	0.55	-	palmitoyl dihydrosphingomyelin (d18:0/16:0)*	Lipid	Dihydrosphingomyelins
-	-	0.52	palmitoyl ethanolamide	Lipid	Endocannabinoid
-	0.56	-	palmitoyl sphingomyelin (d18:1/16:0)	Lipid	Sphingomyelins
-	-	0.53	palmitoylcarnitine (C16)	Lipid	Fatty Acid Metabolism (Acyl Carnitine, Long Chain Saturated)
-	0.50	-	pantothenate (Vitamin B5)	Cofactors and Vitamins	Pantothenate and CoA Metabolism
0.48	-	-	pentadecanoate (15:0)	Lipid	Long Chain Saturated Fatty Acid
-	0.51	-	phenylalanine	Amino Acid	Phenylalanine Metabolism
0.44	-	-	phenyllactate (PLA)	Amino Acid	Phenylalanine Metabolism
0.63	-	-	picolinoylglycine	Lipid	Fatty Acid Metabolism (Acyl Glycine)
0.58	-	-	pregnen-diol disulfate*	Lipid	Pregnenolone Steroids
-	-	-0.53	pregnenediol sulfate (C21H34O5S)*	Lipid	Pregnenolone Steroids

-	-	-0.42	pregnenetriol disulfate*	Lipid	Pregnenolone Steroids
-	-	-0.52	pregnenetriol sulfate*	Lipid	Pregnenolone Steroids
-	0.42	-0.45	pregnenolone sulfate	Lipid	Pregnenolone Steroids
-	-	-0.46	prolylglycine	Peptide	Dipeptide
0.44	-	-	prolylhydroxyproline	Amino Acid	Urea cycle; Arginine and Proline Metabolism
0.45	-	-	pseudouridine	Nucleotide	Pyrimidine Metabolism, Uracil containing
0.77	-	-	pyroglutamine*	Amino Acid	Glutamate Metabolism
0.55	-	-	pyruvate	Carbohydrate	Glycolysis, Gluconeogenesis, and Pyruvate Metabolism
-	-	-0.43	quinolinate	Cofactors and Vitamins	Nicotinate and Nicotinamide Metabolism
0.41	-	-	retinol (Vitamin A)	Cofactors and Vitamins	Vitamin A Metabolism
0.46	-	-	ribitol	Carbohydrate	Pentose Metabolism
0.49	-	-	S-adenosylhomocysteine (SAH)	Amino Acid	Methionine, Cysteine, SAM and Taurine Metabolism
0.49	-	-	sebacate (C10-DC)	Lipid	Fatty Acid, Dicarboxylate

0.42	-	-	sphingomyelin (d17:1/16:0, d18:1/15:0, d16:1/17:0)*	Lipid	Sphingomyelins
-	-	0.56	sphingomyelin (d18:1/14:0, d16:1/16:0)*	Lipid	Sphingomyelins
-	-	0.49	sphingomyelin (d18:1/17:0, d17:1/18:0, d19:1/16:0)	Lipid	Sphingomyelins
-	-	0.59	sphingomyelin (d18:1/18:1, d18:2/18:0)	Lipid	Sphingomyelins
-	-	0.53	sphingomyelin (d18:1/19:0, d19:1/18:0)*	Lipid	Sphingomyelins
-	-	0.55	sphingomyelin (d18:1/20:0, d16:1/22:0)*	Lipid	Sphingomyelins
-	-	0.51	sphingomyelin (d18:1/20:1, d18:2/20:0)*	Lipid	Sphingomyelins
-	-	0.51	sphingomyelin (d18:1/21:0, d17:1/22:0, d16:1/23:0)*	Lipid	Sphingomyelins
-	-	0.57	sphingomyelin (d18:1/22:1, d18:2/22:0, d16:1/24:1)*	Lipid	Sphingomyelins
-	-	0.56	sphingomyelin (d18:2/14:0, d18:1/14:1)*	Lipid	Sphingomyelins
-	-	0.57	sphingomyelin (d18:2/16:0, d18:1/16:1)*	Lipid	Sphingomyelins
-	-	0.56	sphingomyelin (d18:2/18:1)*	Lipid	Sphingomyelins
-	-	0.42	sphingomyelin (d18:2/21:0, d16:2/23:0)*	Lipid	Sphingomyelins

-	-	0.59	sphingomyelin (d18:2/23:0, d18:1/23:1, d17:1/24:1)*	Lipid	Sphingomyelins
-	-	0.66	sphingomyelin (d18:2/23:1)*	Lipid	Sphingomyelins
-	-	0.52	sphingomyelin (d18:2/24:1, d18:1/24:2)*	Lipid	Sphingomyelins
-	-	0.45	stearate (18:0)	Lipid	Long Chain Saturated Fatty Acid
-	0.65	-	stearidonate (18:4n3)	Lipid	Long Chain Polyunsaturated Fatty Acid (n3 and n6)
-	0.59	-	stearoyl sphingomyelin (d18:1/18:0)	Lipid	Sphingomyelins
-	-	0.44	suberate (C8-DC)	Lipid	Fatty Acid, Dicarboxylate
0.44	-	-	succinylcarnitine (C4)	Energy	TCA Cycle
0.57	-	-	sulfate*	Xenobiotics	Chemical
0.53	-	-	tetradecadienedioate (C14:2-DC)*	Lipid	Fatty Acid, Dicarboxylate
-	0.46	-	tetradecadienoate (14:2)*	Lipid	Fatty Acid, Dicarboxylate
-	0.59	-	tetradecanedioate (C14)	Lipid	Fatty Acid, Dicarboxylate
-	0.50	-	thioproline	Xenobiotics	Chemical
-	-	-0.44	tricosanoyl sphingomyelin (d18:1/23:0)*	Lipid	Sphingomyelins
-	-	0.46	tridecenedioate (C13:1-DC)*	Lipid	Fatty Acid, Dicarboxylate
-	0.42	-	tyramine O-sulfate	Amino Acid	Tyrosine Metabolism

0.44	-	-	undecanedioate (C11-DC)	Lipid	Fatty Acid, Dicarboxylate
0.40	-	-	urea	Amino Acid	Urea cycle; Arginine and Proline Metabolism
0.45	-	-	urocortisol glucuronide (4) aka tetrahydrocortisol glucuronide (5)	Lipid	Corticosteroids
-	0.41	-	vanillactate	Amino Acid	Tyrosine Metabolism
0.65	-	-	xanthosine	Nucleotide	Purine Metabolism, (Hypo)Xanthine/Inosine containing

PCA was performed on 874 metabolites followed by a varimax rotation. The first three axes (i.e., principal component) each explained 9.63, 4.79, and 4.69 percent variance of metabolites, respectively. Metabolites with scores ≥ 0.40 or ≤ -0.40 are shown here. Metabolites are ordered alphabetically.

Table S5.3. Characteristics of the analysis sample (n=1,082)

	Mean (SD) or n (%)
Age, years	50.99 (9.14)
Women, n (%)	589 (54.44%)
Systolic blood pressure (SBP), mmHg	126.01 (17.43)
Diastolic blood pressure (DBP), mmHg	80.70 (10.67)
High blood pressure (self-reported or SBP/DBP≥130/80 mmHg)	676 (62.48%)
Body mass index (BMI), kg/m^2	23.78 (3.23)
Estimated glomerular filtration rate (eGFR)[*], mL/min/1.73m^2	76.68 (14.38)
Province, n (%)	
Guangxi	304 (28.10%)
Guizhou	227 (20.98%)
Henan	225 (20.79%)
Hunan	326 (30.13%)
Urbanization index[†]	68.65 (16.52)
Per capita household income[‡], 1000 yuan	19.01 (33.56)
Completed high school education, n (%)	308 (28.47%)
Total energy[§], kcal	2001.31 (645.75)

Animal-source food consumption[§], %kcal	21.70 (13.74)
Sodium consumption[§], mg	4042.60 (2046.17)
Physical activity[‖], METS/week	161.6 (162.1)
Ever smoking, n (%)	459 (42.42%)
Drank alcohol last year, n (%)	321 (29.67%)
Sample with gut microbiota data, n (%)	1,003 (92.70%)
sample with plasma metabolites data, n (%)	434 (43.06%)
Sample with both microbiota and metabolites data, n (%)	355 (32.80%)

[*] eGFR was calculated based on the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation.

[†] Urbanization index encompasses 12 components of urbanization, including population density, sanitation, and health infrastructure.

[‡] Per capita household income was estimated by dividing the total household income by the number of household members.

[§] Dietary intakes were measured by 3-consecutive 24h dietary recalls and household food inventories.

[6] Physical activity was estimated by 7-day physical activity (occupational, transportation, domestic, and leisure activities) recalls in METS.

Table S5.4. Associations between gut microbial within-person diversity (α-diversity) measures with systolic and diastolic blood pressure (SBP and DBP), n=1,003

		SBP				DBP			
		Model 1		Model 2		Model 1		Model 2	
	Mean (SD)	β Coefficient (95% CI)	P value	β Coefficient (95% CI)	P value	β Coefficient (95% CI)	P value	β Coefficient (95% CI)	P value
Shannon index*	2.59 (0.30)	0.90 (-2.52, 4.33)	0.60	0.51 (-2.83, 3.86)	0.76	-0.52 (-2.72, 1.68)	0.64	-0.83 (-2.96, 1.31)	0.45
Richness†	99.14 (27.51)	0.01 (-0.03, 0.05)	0.59	0.01 (-0.03, 0.05)	0.64	-4E-04 (-0.02, 0.02)	0.97	-1E-03 (-0.02, 0.02)	0.91

Model 1 was adjusted for age, sex, provinces, urbanization index (tertiles: ≤58.8, 58.8-78.8, >78.8), per capita household income (tertiles: ≤7.2, 7.2-18, >18), education, total energy intake, animal-source food consumption, sodium consumption, physical activity (tertiles: ≤75, 75-192, >192), smoking, alcohol intake, and estimated glomerular filtration rate (eGFR). Model 2 was additionally adjusted for BMI.

*Shannon index at genus level was calculated using $-\sum p_i ln p_i$, where p_i is the proportional abundance of genera i.

†Richness measured the number of distinct genera per subject with rarefication.

Table S5.5. Association between gut microbial principal coordinates analysis (PCoA) axes with systolic and diastolic blood pressure (SBP and DBP, mmHg), Coefficient (95% CI)

	Taxa contributing to each axis		**Eigen-value**	**Variance explained**	**SBP**		**DBP**	
					Model 1	**Model 2**	**Model 1**	**Model 2**
MDS1	*Prevotella*	*Bacteroides*	4.91	8.61%	18.08	16.97	3.27	2.53
	Ralstonia	*Unassigned*			(-15.40,	(-15.70,	(-18.27,	(18.33,
	S24-7	*Comamonadaceae*			51.56)	49.64)	24.80)	23.40)
	Delftia	*Rikenellaceae*						
	Parabacteroides	*RF39*						
MDS 2	*Ralstonia*	*Akkermansia*	3.18	5.58%	6.01	4.80	-4.89	-5.67
	Delftia	*Comamonadaceae*			(-27.13,	(-27.57,	(-26.20,	(-26.34,
	Bacteroides	*Prevotella*			39.16)	37.18)	16.41)	14.99)
	Lachnospira	*Lactococcus*						
	Faecalibacterium	*Phascolarctobacterium*						
MDS3	*Lactococcus*	*YS2*	2.02	3.54%	15.45	5.81	3.96	-2.97
	Eggerthella	*Bifidobacterium*			(-30.06,	(-38.69,	(-25.31,	(-31.37,
	Catenibacterium	*Peptococcus*			60.97)	50.30)	33.22)	25.44)
	Leuconostocaceae	*Adlercreutzia*						

	Pseudoramibacter_ Eubacterium	*Fusobacterium*						
MDS4	*Adlercreutzia* *Enterobacteriaceae_other* *Leuconostocaceae* *Serratia* *Enterobacteriaceae*	*Rothia* *Fusobacterium* *Coprococcus* *Eggerthella* *Ralstonia*	1.80	3.12%	52.94 (15.61, 90.27)**	38.90 (2.19, 75.61) *	16.60 (-7.47, 40.67)	6.47 (-17.01, 29.95)

MDS, multidimensional scaling. PCo were derived from principal coordinate analysis of 1,008 taxa based on Bray-Curtis distance matrix in microbiota analysis sample (n=1,008). Coefficient indicates SBP and DBP (mmHg) associated with each 1 unit increase in microbiota pattern score. Contributing taxa include taxa with the top 10 highest absolute scores for the respective PCo. Taxa are vertically ordered by their absolute values of scores (descending). A full list of scores is shown in Table S5.1. Model 1 was adjusted for age, sex, provinces, urbanization index (tertiles: ≤58.8, 58.8-78.8, >78.8), per capita household income (tertiles: ≤7.2, 7.2-18, >18), education, total energy intake, animal-source food consumption, sodium consumption, physical activity (tertiles: ≤75, 75-192, >192), smoking, alcohol intake, and estimated glomerular filtration rate (eGFR). Model 2 was additionally adjusted for BMI. "*" indicates p-value < 0.05; "**" indicates p-value<0.01

Figure S5.1. Sample flow chart.

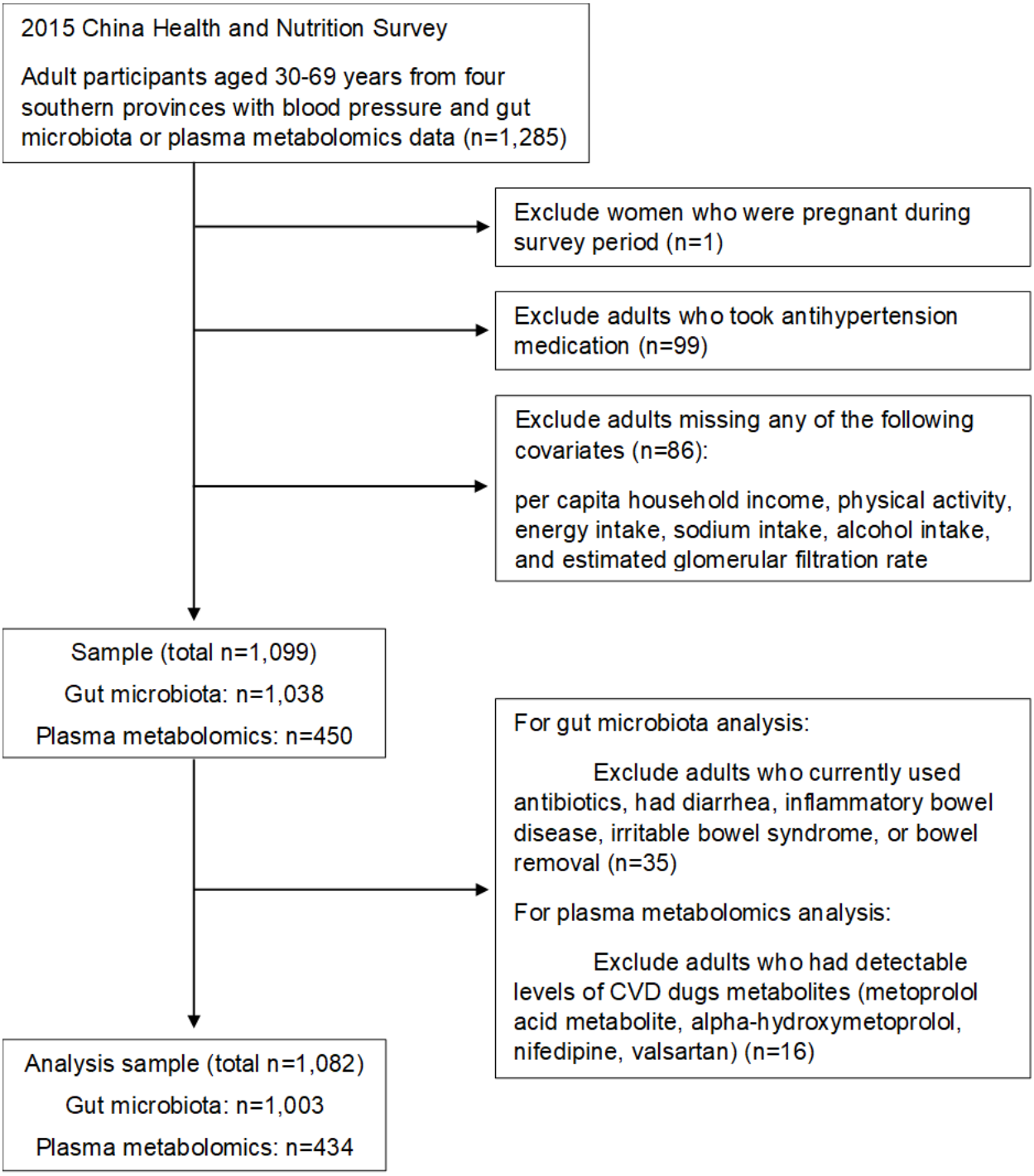
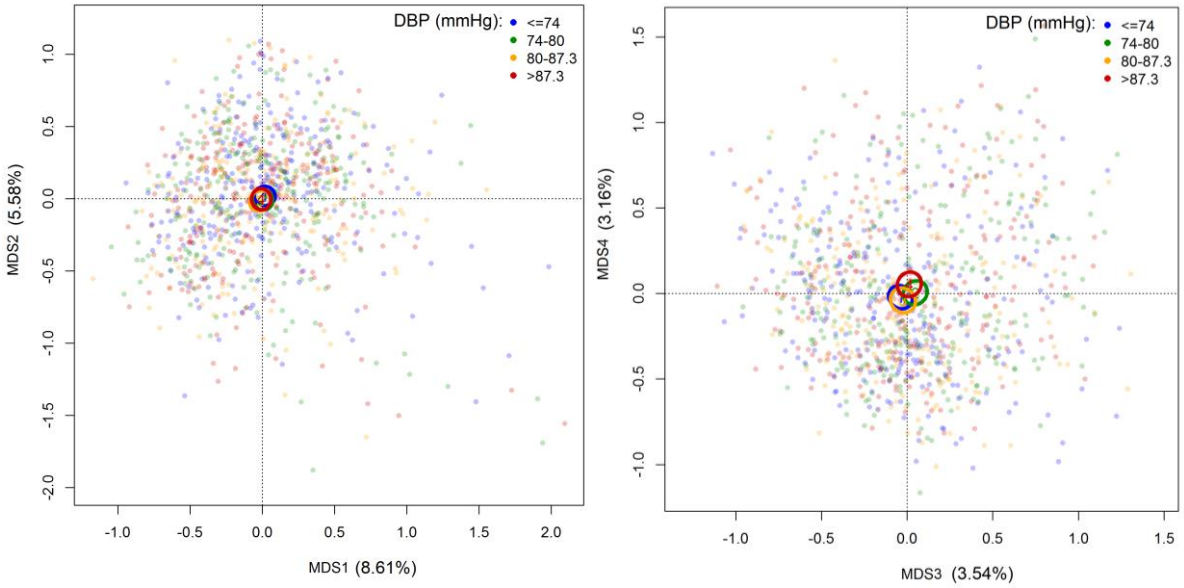


Figure S5.2. Microbial between-person diversity (β-diversity) assessed using principal coordinate analysis (PCoA) by diastolic blood pressure (SBP). MDS, multidivisional scaling (i.e., PCoA axis). For better visualization of the separation, DBP was categorized to quartiles. Plot legend shows the color and range of DBP for each quartile. Centroids illustrate the 95% CI for the mean location of each DBP quartile. In permutational multivariate analysis of variance (PERMANOVA), DBP (continuous) had R^2 of 1.42% and p-value of 0.026 in Model 1 and 0.046 in Model 2. Model 1 was adjusted for age, sex, provinces, urbanization index (tertiles: ≤64.2, 64.2-81.5, >81.5), per capita household income (tertiles: ≤10, 10-21.6, >21.6), education, total energy intake, animal-source food consumption, sodium consumption, physical activity (tertiles: ≤57.4, 57.4-152, >152), smoking, alcohol intake, and estimated glomerular filtration rate. Model 2 was additionally adjusted for BMI.



CHAPTER 6. CIRCULATING SHORT-CHAIN FATTY ACIDS ARE POSITIVELY ASSOCIATED WITH ADIPOSITY MEASURES IN CHINESE ADULTS

Overview

Epidemiological studies suggest a positive association between obesity and fecal short-chain fatty acids (SCFAs) produced by microbial fermentation of dietary carbohydrates, while animal models suggest increased energy harvest through colonic SCFA production in obesity. However, there is a lack of human population-based studies with dietary intake data, plasma SCFAs, gut microbial, and anthropometric data.

In 490 Chinese adults aged 30-68y, we examined the associations between key plasma SCFAs (butyrate/isobutyrate, isovalerate, and valerate measured by nontargeted plasma metabolomics) with BMI using multivariable-adjusted linear regression. We then assessed whether overweight (BMI≥24kg/m^2) modified the association between dietary-precursors of SCFAs (insoluble fiber, total carbohydrates, and high-fiber foods) with plasma SCFAs. In a sub-sample (n=209) with gut metagenome data, we examined the association between gut microbial SCFA-producers with BMI.

We found positive associations between butyrate/isobutyrate and BMI (p-value<0.05). The associations between insoluble fiber and butyrate/isobutyrate differed by overweight (p-value<0.10). There was no statistical evidence for an association between microbial SCFA-producers and BMI. In sum, plasma SCFAs were positively associated with BMI and that the colonic fermentation of fiber may differ for adults with versus without overweight.

Introduction

Overall and central obesity are major risk factors for a wide range of chronic diseases, including cardiometabolic diseases [1-3]. As the prevalence of obesity has increased dramatically over the past decades around the world [1], many studies have been conducted to identify the biological determinants of obesity. Recent evidence has shown that the gut microbiota and microbiota-mediated metabolites like short-chain fatty acids (SCFAs) influence diet-induced obesity [4,5]. SCFAs like butyrate are major products of microbiota fermentation of dietary carbohydrates, especially soluble fiber and resistant starch [4]. In human studies, fiber-rich diets and Mediterranean diets have been shown to be positively associated with weight loss [6,7] and increased serum [8] and fecal SCFAs [9], respectively.

However, studies have yielded incongruent results for the SCFA-obesity association, which involves various factors like diet and gut microbiota. Whereas several studies have demonstrated that dietary SCFA supplementation may be beneficial to weight loss through appetite regulation [10,11] and increases in lipid oxidation and energy expenditure [12], others have suggested that SCFA production may promote obesity [5,13-15] through pathways including *de novo* lipogenesis [13] and energy harvesting from diet by gut microbiota [5]. For example, a randomized, controlled study showed that colonic delivery of SCFA propionate (i.e., oral supplementation of inulin-propionate ester) reduced energy intake, weight gain, and intra-abdominal fat accretion in overweight adults [10]. In contrast, in a study of obesity-prone mice fed a macronutrient-matched and isoenergetic high-fat diet, Isken et al. suggested that colonic SCFA production potentially outweighed the beneficial effects of soluble fiber supplementation on diet-induced obesity via contribution to increased digested energy [15]. Additionally, mouse model and in vitro assays suggest that the gut microbiome of mice and humans with (versus without) obesity had increased capacity to harvest energy through colonic fermentation of dietary carbohydrates and SCFA production [5,13].

There has been a lack of large population-based studies with large variation in dietary intake and gut microbiota composition to examine the incongruent experimental results in free-living people. Although a few community-based studies and case-control studies of fecal SCFAs in Western human populations

support positive associations between SCFAs with overall body mass and central adiposity [16-20], few population-based studies have examined plasma SCFAs, which, in contrast to fecal SFCAs, may better represent the fraction of SCFAs that enter the host blood stream as a potential source of energy [21]. Therefore, we aimed to investigate the associations between plasma SCFAs with two adiposity measures, body mass index (BMI) and waist-to-height ratio (WHtR) in a socio-demographically diverse cohort of Chinese adults consuming a range of traditional and Western diets. We also assessed whether overweight and abdominal obesity modified the association between dietary precursors of SCFAs and plasma SCFAs. In a sub-sample with gut metagenome data, we examined the association between gut microbial SCFA producers with BMI and WHtR.

Methods

Study sample

We used nontargeted plasma metabolomics data from the 2015 China Health and Nutrition Survey (CHNS). Adults with anthropometry and diet data were ligible for the current study (n=500). We excluded participants if they were pregnant (n=1) or had missing covariates (n=9), resulting in an analysis sample of 490 adults, among which a subset of 209 adults also had gut metagenome data (Figure S1).

Measures

We detected three SCFAs among the 1,108 matched compounds: butyrate/isobutyrate, valerate, and isovalerate. Metabolon rescaled the raw area count of each metabolite within each run-day to a median of 1 to correct for differences in instrument inter-day tuning and imputed values below detection limits with the minimum. We $\log_2$ transformed the abundance of these three SCFAs. As the total SCFAs was also of interest, we $\log_2$ transformed the sum of the total abundance for all three SCFAs.

We calculated BMI as weight divided by squared height (kg/m^2) and WHtR as waist circumference divided by height. We defined overweight as BMI ≥24 kg/m^2 and abdominal obesity as WHtR ≥0.5, according to the optimal cut-off points to indicate cardiovascular diseases risk in Chinese adults [25,26].

We included three measures of dietary precursors of SCFAs: insoluble fiber, total carbohydrates, and an *a priori* high-fiber food group (Table S1) consisting of: whole grains (e.g. millet), legumes (e.g. soybean curd), starchy roots (e.g. potato), vegetables (e.g. cabbage), mushrooms/seaweeds (e.g. Shitake mushroom), fruits (e.g. apple), nuts/seeds (e.g. walnut). We grouped insoluble fiber, carbohydrate, and high-fiber foods by tertiles to limit the influence of extreme consumers, allow for non-linearity of relationships, and preserve statistical power. In analysis of individual foods, we categorized those consumed by more than and less than 50% of the sample by median and any/no intake, respectively.

For measures of gut microbial SCFA producers, we selected 56 potential SCFA-producing microbiota species based on literature search (Table S2) and calculated the total counts of the 56 selected species. We normalized and $\log_{10}$ transformed the raw counts of each species and the total counts of selected species [29]. For analysis of specific species, we dichotomized 27 rare species that present in less than 25% of the sample to yes/no detected in the sample.

Statistical analysis

We presented continuous variables as mean (SD) and categorical variables as number (proportion). We compared the characteristics by overweight and abdominal obesity using t-test for continuous variables and chi-square test for categorical variables.

To determine the associations between plasma SCFAs with BMI and WHtR, we used a linear regression adjusting for the following covariates as guided by literature [4,31-36]: age, sex, batch run,

province, urbanization, education, income, energy intake, insoluble fiber intake, physical activity, smoking, and alcohol intake.

To investigate whether overweight and abdominal obesity modified the association between dietary precursors of SCFAs (insoluble fiber, total carbohydrates and high-fiber foods) with plasma SCFAs, we assessed the interaction of each of these dietary precursors with overweight and abdominal obesity in linear regression models of plasma SFCAs using a Wald test at a nominal significance level of p-value <0.10. In an exploratory analysis, we examined the interaction of each individual types of foods included in the high-fiber food group with overweight and abdominal obesity in linear regression models of plasma SCFAs. Then, in the sub-sample also containing gut metagenome data, we examined (1) the association between the overall microbial SCFA producers (i.e., SCFA-producing species) with BMI and WHtR using permutational multivariate analysis of variance (PERMANOVA) based on Bray-Curtis distance with 999 permutations [37] and (2) the association between the total relative abundance of all microbial SCFA producers with BMI and WHtR using linear regression. In an exploratory analysis, we examined each individual microbial SCFA producer using linear regression. For sensitivity, we tested the associations between microbial SCFA producers and plasma SCFAs. All analyses were adjusted for covariates described above.

We conducted sensitivity analysis by excluding participants who took antibiotics, pre/probiotics, or had diarrhea, irritable bowel syndrome (IBS), or inflammatory bowel disease (IBD), because these factors may affect the gut microbiome and therefore influence SCFA production and absorption. After applying these exclusion criteria, 462 adults remained in the analysis sample, among which 192 had gut metagenome data. We performed all statistical analyses in R 3.6.0 (http://www.r-project.org).

Results

The prevalence of overweight and abdominal obesity were 48.8% and 65.2%, respectively, in our analysis sample (Table 1). Adults with overweight were not different from those without overweight in terms of age, sex, plasma SCFAs, province, urbanization, education, income, diet, physical activity, smoking, and alcohol intake. Adults with abdominal obesity were older, less physically active, and had higher abundance of plasma butyrate/isobutyrate, isovalerate, and total SCFAs than those without abdominal obesity.

We first examined the associations between plasma SCFAs and BMI and WHtR and found that butyrate/isobutyrate was positively associated with BMI and WHtR, and isovalerate and total SCFAs were positively associated with WHtR (p-value<0.05, Table 2). For example, a fold increase of butyrate/isobutyrate was associated with a 0.40 and a 0.01 unit increase in BMI (kg/m^2) and WHtR, respectively.

Then, in models of plasma SFCAs, we tested the interaction between dietary precursors of SCFAs with overweight (Figure 1, Table S3) and abdominal obesity (Figure 2, Table S4). We observed effect modification of the association (interaction p-value<0.10) between insoluble fiber with butyrate/isobutyrate by overweight; between insoluble fiber with butyrate/isobutyrate and total SCFAs by abdominal obesity; and between carbohydrate with valerate by abdominal obesity. Whereas the model estimated abundance of valerate was lower at high versus low carbohydrate consumption in people without abdominal obesity, valerate abundance was slightly higher at high versus low carbohydrate consumption in people with abdominal obesity. Moreover, when consuming middle level of insoluble fiber, adults with abdominal obesity had higher abundance of butyrate/isobutyrate and total SCFAs than those without abdominal obesity.

In an exploratory analysis, we examined individual types of foods included in the high-fiber food group and observed effect modification (interaction p-value<0.10) of the associations between legumes

and fruits with valerate by overweight (Table S5); between whole grains and nuts/seeds with butyrate/isobutyrate, valerate and total SCFAs by abdominal obesity (Table S6); and between fruits with valerate and total SCFAs by abdominal obesity. In general, consuming more of these fiber-rich foods tended to be associated with lower SCFAs in adults with abdominal obesity.

Last, in the sub-sample with gut metagenome data, we tested whether adults with higher BMI or WHtR had higher relative abundance of gut microbial SCFA producers. We found little statistical evidence of association between the overall microbial SCFA producers with BMI (PERMANOVA R^2 = 0.008, Table 3) or WHtR (PERMANOVA R^2 = 0.005). There was no statistical evidence of association between the total relative abundance of microbial SCFA producers with BMI and WHtR either (p-value>0.500). In an exploratory analysis examining the specific microbial SCFA producers, we found a few species that were associated with BMI and WHtR (Table S7) at p-value<0.05. For example, *Eubacterium hallii* and *Eubacterium rectale* were positively associated with BMI and WHtR. For sensitivity, we tested whether those species were associated with plasma SCFAs. We found no association between the overall (PERMANOVA R^2 ranged 0.002-0.004, Table S8) or the total relative abundance (p-value>0.100) of microbial SCFA producers with plasma SCFAs, though a few individual species were associated with butyrate/isobutyrate, valerate, isovalerate, and total plasma SCFAs (Table S9) at p-value<0.05, including *Faecalibacterium prausnitzii*.

In sensitivity analysis that restricted the sample to those who did not take antibiotics, pre/probiotics, or had diarrhea, IBS, or IBD, though the statistical significance reduced a little due to decreased sample size, the patterns of associations and parameter estimates remained similar for the associations between plasma SCFAs and BMI and WHtR (Table S10); interactions of dietary precursors of SCFAs with overweight (Figure S2, Table S11) and abdominal obesity in linear models of plasma SCFAs (Figure S2, Table S12); and associations of the overall and total relative abundance of microbial SCFA producers with BMI and WHtR (Table S13).

Discussion

In our population-based cohort of Chinese adults, we found positive associations between plasma SCFAs and two adiposity measures, BMI and WHtR, independent of sociodemographic and behavioral factors, including urbanization, diet, and physical activity. Butyrate/isobutyrate was positively associated with BMI; and butyrate/isobutyrate, isovalerate, and total SCFAs were positively associated with WHtR. Our results provide insights into the potential role of SCFAs in the etiology of obesity and abdominal obesity.

Several cross-sectional studies have reported positive associations between fecal SCFAs levels and obesity [16-20]. For example, a community-based study of 441 Colombian adults aged 18-62 years demonstrated that higher fecal butyrate, acetate, propionate and total SCFAs were associated with BMI, body fat, and waist circumference [17]. One hypothesis is that gut microbial dysbiosis in obesity may lead to less efficient SCFA absorption and therefore more SCFA excretion [17,38]. However, Rahat-Rozenbloom et al. [19] suggested that in 22 Canadian individuals aged >17 years, higher fecal acetate, butyrate, and total SCFAs in people with (versus without) overweight was not due to differences in diet or SCFAs absorption measured by rectal dialysis bag method. Our findings of positive associations between plasma SCFAs and adiposity measures may support their results that higher SCFA excretion in higher body mass was not due to reduced SCFA absorption [19], though studies using both circulating and fecal SCFAs are needed to fully elucidate this hypothesis. Moreover, our findings show no difference in intakes of dietary precursors of SCFAs by overweight and abdominal obesity, but potential effect medication of associations between dietary precursors of SFCAs (e.g., insoluble fiber) and plasma SCFAs (i.e., butyrate/isobutyrate) by overweight and abdominal obesity, indicating that colonic fermentation of dietary precursors of SCFAs may differ for people with and without overweight, thereby leading to different abundance of plasma SCFAs, Though we focused on well-established dietary precursors like fiber and carbohydrates, our results in butyrate/isobutyrate and isovalerate suggest that higher protein catabolism may be associated with higher adiposity, given that isobutyrate and isovalerate are major fermentation

products of amino acids valine and leucine, respectively [39]. These branched-chain SCFAs function similarly to straight-chain SCFAs (e.g. butyrate) when modulating glucose and lipid metabolism [40].

Additionally, our results are consistent with Goffredo et al. [13], which showed that plasma concentrations of acetate, propionate, and butyrate were positively associated with body fat percentage and BMI changes in 84 children and adolescents. Conversely, in 12 normal- and over-weight adults aged 18-65 years, Boets et al. [41] found that the rate of appearance of plasma propionate and butyrate measured by stable isotope dilution was lower in subjects with higher BMI. In 18 women who were obese, Layden et al. [42] found a negative association between serum acetate and visceral adipose tissue (though not BMI). These two studies had much smaller samples with less variation in BMI than our study.

A potential reason for the positive relationship between SCFAs and obesity is that the gut microbiota of individuals with obesity may have higher capacity to harvest energy through SCFA production [5,13]. In fact, SCFAs are estimated to add about 10% of extra daily energy intake to adults eating westernized diet [14]. Goffredo et al. [13] found that plasma SCFAs were associated with hepatic de novo lipogenesis and the gut microbiota of adolescents with obesity compared to their lean counterparts had higher capacity to ferment the same amount of fructose in vitro. Additionally, Yang et al. showed that the fecal microbiota of people with (versus without) obesity produced more propionate in response to in vitro fermentation of cereal grains [43]. Similarly, we found that carbohydrate consumption tended to be negatively associated with plasma valerate in adults without abdominal obesity, but tended to be positively associated with plasma valerate in adults with abdominal obesity, indicating that adults with higher abdominal adiposity may have higher potential for carbohydrate fermentation when consuming a carbohydrate-rich diet. The fact that fruits, whole grains, and nuts/seeds consumption tended to be negatively associated with SCFAs in adults with abdominal obesity in our study suggests that consuming these healthy fiber-rich foods may help decrease energy harvesting via SCFA production in adults with excess abdominal adiposity. Indeed, a double-blinded, randomized-controlled trial showed that fecal

acetate, propionate and total SCFAs were lower in prebiotic dietary fibers-treated group than placebo group after treatment [44]. We also found different patterns of associations between different dietary precursors of SCFAs and plasma SCFAs. This may be because different gut microbiota compositions respond differently to different sources of fermentable fibers [43,45].

Nevertheless, not all epidemiological studies that support positive associations between SCFAs and adiposity measures also suggest the potential involvement of gut microbiota [13,16,17,19,20]. Whereas a recent meta-analysis of case-control studies found no evidence of association between obesity and phyla richness [16], other studies found that higher fecal SCFAs was associated with lower microbiota diversity and higher gut permeability, Firmicutes/Bacteriodetes ratio and cardiometabolic disease-associated taxa [17,19]. While a study of 96 adolescent girls found no association between microbiota phyla abundance and fecal SCFAs [20], a study 84 adolescent boys and girls found a positive association between plasma SCFAs and obesity-related microbiota, including *Faecalibacterium*, *Streptococcus* and *Actinomyces* [13]. Although our sub-sample analysis using gut metagenome data showed a few positive associations between microbial producers of SCFAs with BMI, WHtR, and plasma SCFAs, like *Eubacterium hallii* and *Eubacterium rectale*, none of which reached statistical significance after considering multiple testing (Bonferroni corrected p-value threshold =0.05 / 56 species / 6 outcomes = 1.5E-4). The relationships between gut microbiota with SCFAs production and host adiposity are complicated given the convoluted metabolic pathways and bacterial cross-feeding interactions [46]. It is possible that we lacked statistical power to detect these associations, given that our sample comprised free-living people from a range of rural and urban communities, providing more diversity in environments and behaviors, than previous studies that were conducted in a single city or neighborhood [17,19].

The strengths of our study include the well-characterized, population-based cohort, high-quality dietary data from three-consecutive 24-h recalls and household food inventories, and rich host factors collected from standardized protocols, allowing us to control for a wide range of potential confounders. Additionally, to our knowledge, we are the first study to examine the associations between plasma

SCFAs, gut metagenome, and adiposity measures in population-based adult cohort and our study is relatively larger than previous studies studying fecal SCFAs [16-20]. However, it is possible that we lacked statistical power to test potential interactions between dietary precursors of SCFAs with overweight and abdominal obesity in models of plasma SCFAs. After considering multiple testing (3 precursors X 4 SCFAs X 2 adiposity measures =24 tests), none of the interactions reached statistical significance (Bonferroni corrected p-value threshold =0.10/24 =4.2E-03). Other limitations of our study include the (1) cross-sectional design; (2) potential measurement error in diet and lack of detailed information on fiber types to distinguish fermentable versus non-fermentable fiber; (3) selected SCFA-producing species may not comprehensively capture the SCFA-producing ability of the gut microbiota; (4) nontargeted metabolomics could not provide the concentrations for SCFAs and could not detect two major SCFAs (acetate and propionate) due to small molecular sizes. The fraction of acetate:propionate:butyrate in portal system is approximately 69:23:8 [47]. Since butyrate could be undetectable in circulation because of rapid usage [48], we could not exclude the possibility that we underestimated the abundance of plasma butyrate and other SCFAs.

Conclusion

Our study in a population-based cohort of Chinese adults suggests that plasma SCFAs may be positively associated with overall body mass and abdominal girth measured by BMI and WHtR, respectively, providing insights into the possible role of SCFAs in obesity etiology. Our results also suggest that individuals with higher adiposity may potentially have higher capacity to produce SCFAs from dietary carbohydrate and insoluble fiber than lean individuals, though we were unable to confirm whether it is due to differences in relative abundance of SCFA-producing gut microbiota. Further studies are needed to confirm our results, determine the causal relationship between SCFAs and adiposity, and clarity whether the gut microbiota in people with higher adiposity harvest more energy from diet than those with lower adiposity through SCFAs production.

Tables and figures

Table 6.1. Characteristics of the metabolomics analysis sample by body mass index (BMI) and waist-to-height ratio (WHtR) groups

	Overweight[1]		**Abdominal obesity[1]**	
	Without	**With**	**Without**	**With**
n (%)	251 (51.2%)	239 (48.8%)	170 (34.8%)	318 (65.2%)
Age, years	52.3 (9.0)	52.2 (9.1)	51.0 (8.8)	52.9 (9.1)*
Women, n (%)	155 (61.8%)	135 (56.5%)	100 (58.8%)	189 (59.4%)
Body mass index (BMI), kg/m^2	21.8 (1.7)	26.6 (2.0)***	21.1 (2.0)	25.5 (2.7)***
Waist-to-height-ratio (WHtR)	0.5 (0.05)	0.6 (0.05)***	0.5 (0.03)	0.6 (0.04)***
Butyrate/isobutyrate[2]	-0.05 (0.7)	-0.001 (0.8)	-0.2 (0.7)	0.04 (0.7)**
Valerate[2]	-1.1 (1.8)	-1.1 (1.7)	-1.1 (1.7)	-1.0 (1.8)
Isovalerate[2]	0.07 (0.9)	0.1 (1.0)	-0.03 (0.9)	0.2 (0.9)*
Total short-chain fatty acids (SCFAs)[2]	1.5 (0.8)	1.5 (0.8)	1.4 (0.8)	1.6 (0.7)**
Hunan province	159 (63.3%)	145 (60.7%)	102 (60.0%)	200 (62.9%)
Urbanization index, n (%)[3]				
Low (39.2 – 64.2)	89 (35.5%)	84 (35.1%)	63 (37.1%)	110 (34.3%)
Middle (64.2 – 81.5)	77 (30.7%)	84 (35.1%)	51 (30.0%)	109 (34.3%)
High (81.5 – 99.6)	85 (33.9%)	71 (29.7%)	56 (32.9%)	99 (31.1%)

Completed high school education	85 (33.9%)	65 (27.2%)	52 (30.6%)	96 (30.2%)
Per capita household income, n (%))[4]				
Low (0 – 10k yuan)	89 (35.5%)	74 (31.0%)	60 (35.3%)	102 (32.1%)
Middle (10 – 22.1k yuan)	85 (33.9%)	79 (33.1%)	65 (38.2%)	99 (31.1%)
High (22.1 – 468k yuan)	77 (30.7%)	86 (36.0%)	45 (26.5%)	117 (36.8%)
Total energy, 1000kcal[5]	1.9 (0.6)	1.9 (0.7)	1.9 (0.6)	1.9 (0.7)
Insoluble fiber intake, n (%)[5]				
Low (1.5 – 8.2g)	77 (30.7%)	87 (36.4%)	58 (34.1%)	105 (33.0%)
Middle (8.2 – 12.5g)	88 (35.1%)	75 (31.4%)	57 (33.5%)	105 (33.0%)
High (12.5 – 69.7g)	86 (34.3%)	77 (32.2%)	55 (32.4%)	108 (34.0%)
Carbohydrate intake, n (%)[5]				
Low (65.2 – 172g)	83 (33.1%)	81 (33.9%)	55 (32.4%)	108 (34.0%)
Middle (172 – 248g)	87 (34.7%)	76 (31.8%)	61 (35.9%)	101 (31.8%)
High (248 – 649g)	81 (32.3%)	82 (34.3%)	54 (31.8%)	109 (34.3%)
High-fiber foods, n (%)[5]				
Low (0 – 344g)	87 (34.7%)	77 (32.2%)	61 (35.9%)	102 (32.1%)
Middle (344 – 482g)	85 (33.9%)	78 (32.6%)	55 (32.4%)	108 (34.0%)
High (482 – 1200g)	79 (31.5%)	84 (35.1%)	54 (31.8%)	108 (34.0%)
Physical activity, n (%)[6]				*
Low (0 – 50 METS/wk),	76 (30.3%)	84 (35.1%)	43 (25.3%)	117 (36.8%)
Middle (50 – 147 METS/wk)	82 (32.7%)	83 (34.7%)	58 (34.1%)	106 (33.3%)
High (147 – 1390 METS/wk)	93 (37.1%)	72 (30.1%)	69 (40.6%)	95 (29.9%)
Ever smoking, n (%)	100 (39.8%)	93 (38.9%)	74 (43.5%)	118 (37.1%)
Drank alcohol last year, n (%)	62 (24.7%)	64 (26.8%)	38 (22.4%)	87 (27.4%)

Continuous variables [mean (SD)] were tested by t test and categorical variables [n (%)] were tested by chi-square test. *, p-value<0.05; **, p-value<0.01; ***, p-value<0.001 when comparing high versus normal BMI and WHtR.

[1] Overweight: BMI ≥24 kg/m^2; abdominal obesity: waist-to-height ratio ≥0.5.

[2] Plasma SCFAs were measured by nontargeted metabolomics and the total SCFAs was estimated by the sum of the three identified SCFAs. The abundance was $\log_2$ transformed.

[3] Urbanization index encompasses 12 dimensions of urbanization, including population density, health infrastructure, and transportation. Urbanization index was categorized by tertiles to represent low, middle, and high levels of urbanization.

[4] Per capita household income was estimated by dividing the household income by the number of household members. Per capita household income was categorized by tertiles to represent low, middle, and high levels of income.

[5] Dietary intakes were measured by 3-consecutive 24h dietary recalls and household food inventories. The intake of high-fiber foods was calculated as the sum of whole grains, legumes, starchy roots, vegetables, mushrooms/seaweeds, fruits, nuts/seeds. Insoluble fiber, carbohydrate, and high-fiber food score were categorized by tertiles to represent low, middle, and high intakes.

[6] Physical activity was estimated by 7-day physical activity recalls in METS and was categorized by tertiles to represent low, middle, and high levels of physical activity.

Table 6.2. The associations between plasma short-chain fatty acids (SCFAs) and body mass index (BMI) and waist-to-height ratio (WHtR)

		BMI (n=490)		**WHtR (n=488)**	
	Mean (SD)	**β (95% confidence interval)**	**p-value**	**β (95% confidence interval)**	**p-value**
Butyrate/isobutyrate	-0.03 (0.75)	0.40 (0.01, 0.78)	0.04	0.01 (4E-03, 0.02)	3E-03
Valerate	-1.09 (1.75)	-0.01 (-0.17, 0.16)	0.93	1E-03 (-2E-03, 4E-03)	0.48
Isovalerate	0.09 (0.94)	0.20 (-0.10, 0.52)	0.18	0.01 (3E-04, 0.01)	0.04
Total SCFAs	1.61 (0.76)	0.29 (-0.09, 0.66)	0.14	0.01 (3E-03, 0.02)	0.01

The mean (SD) for BMI (kg/m^2) and WHtR was 24.01 (3.18) and 0.52 (0.06), respectively. Because the SCFAs relative abundance were log_2 transformed, the linear model coefficients are interpreted as units of BMI and WHtR associated with a fold increase in SCFAs. Model was adjusted for age, sex, batch, province, urbanization, income, education, physical activity, total energy intake, fiber intake, alcohol, and ever smoking.

Table 6.3. The associations between the overall and the total relative abundance of 56 gut microbial short-chain fatty acid (SCFA) producers with body mass index (BMI) and waist-to-height-ratio (WHtR)

			Overall[1]		**Total[2]**	
	n	**Mean (SD)**	**R^2**	**P-value**	**β (95% CI)**	**P-value**
BMI	209	24.35 (3.21)	0.008	0.05	-0.04 (-1.7, 1.61)	0.96
WHtR	208	0.53 (0.06)	0.005	0.30	0.00 (-0.03, 0.03)	0.78

The 56 SCFA-producing species were selected from literature and the full list with references is in Table S2. The raw counts of each species and the total counts of the 56 species were normalized and $\log_{10}$ transformed [104]. Model was adjusted for age, sex, batch, province, urbanization, income, education, total energy, insoluble fiber, physical activity, smoking, and alcohol intake.

[1] R^2 and p-value were calculated using permutational multivariate analysis of variance (PERMANOVA) of all 56 species.

[2] Linear regression was performed on the total relative abundance of the 56 species.

Figure 6.1. The associations between dietary precursors of short-chain fatty acids (SCFAs) and plasma SCFAs by overweight. Overweight: BMI⩾24 kg/m^2. Vertical axes represent model predicted (marginal means) plasma SCFAs abundance. Dietary intakes of were categorized by tertiles. Linear model was adjusted for age, sex, batch run, province, urbanization, income, education, physical activity, total energy intake, insoluble fiber, alcohol, and ever smoking. P-value for interaction between each dietary precusors and overweight was derived using a Wald test. P-value>0.05 for all comparisons of plasma SCFA abundance at a given level of a dietary precursor by overweight

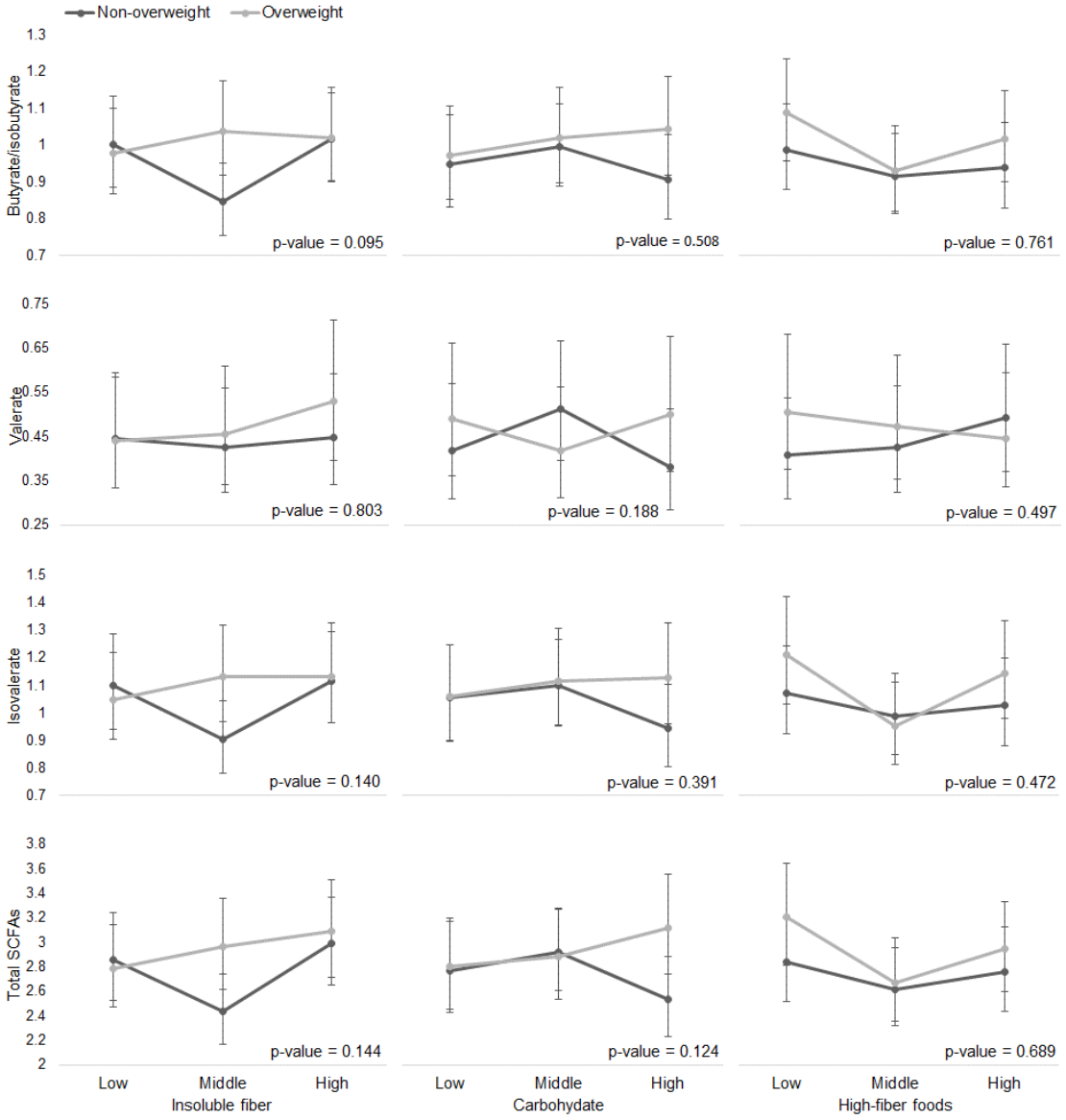


Figure 6.2. The associations between dietary precursors of short-chain fatty acids (SCFAs) and plasma SCFAs by abdominal obesity. Abdominal obesity: waist-to-height ratio⩾0.5. Vertical axes represent model predicted (marginal means) SCFAs abundance. Dietary intakes were categorized by tertiles. Linear model was adjusted for age, sex, batch run, province, urbanization, income, education, physical activity, total energy intake, insoluble fiber intake, alcohol, and ever smoking. P-value for the interaction between each dietary precursor of SCFAs and AOB was derived using a Wald test. *, p-value<0.5; **, p-value <0.01 for comparisons of plasma SCFAs abundance at a given level of dietary precursor by abdominal obesity.

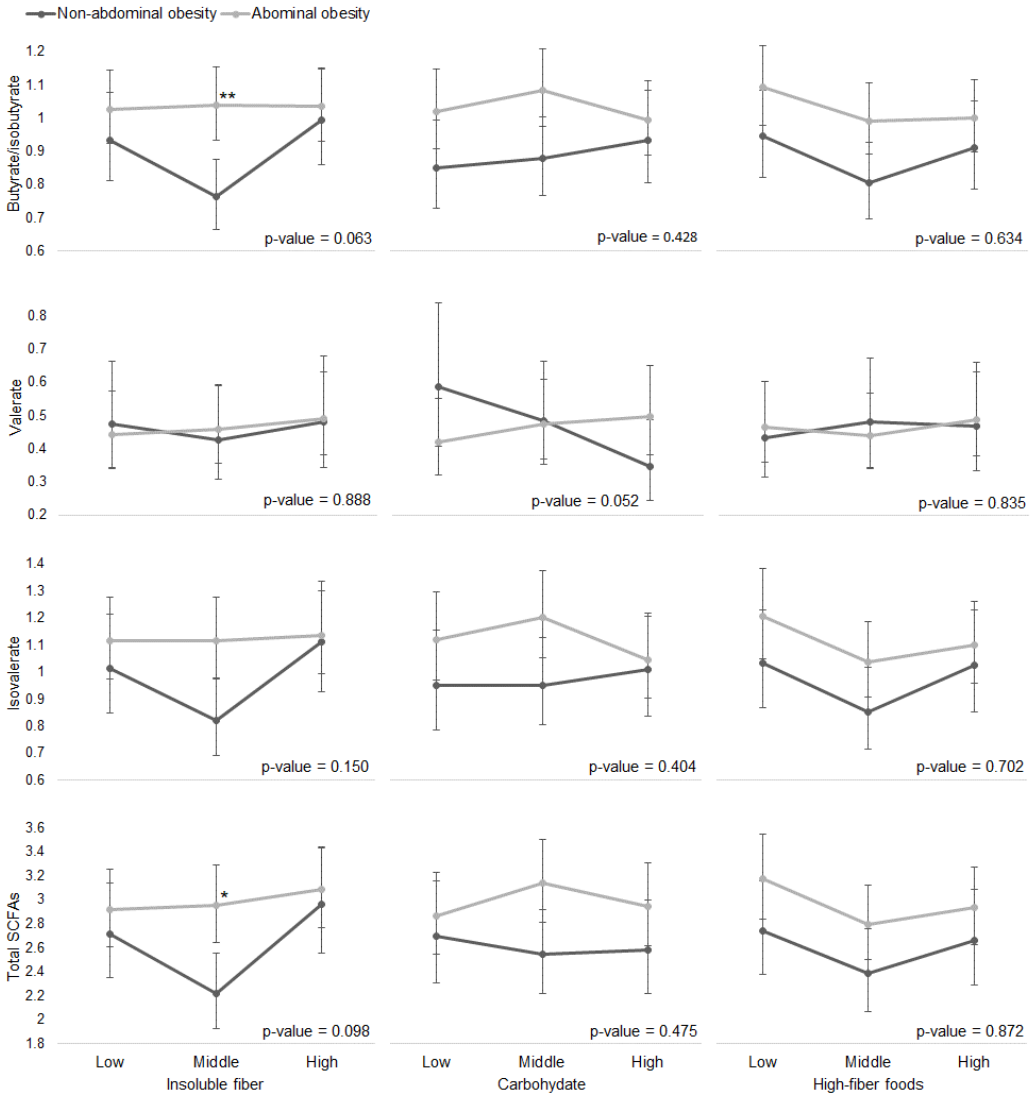


Table S6.1. Foods included in the high-fiber food group

Foods	Examples
Whole grains	Buckwheat flour, foxtail millet, pea starch noodle, tartarian buckwheat flour, yellow corn (grain, grits, flour)
Legumes	Broad bean, soybean curd (soft, semisoft, slab), red bean
Starchy roots	Lotus root, sweet potato, taro, whit potato, winged yam, yam, yam bean
Vegetables	Amaranth, asparagus, bamboo shoot, bitter melon, bok choy, broccoli, cabbage, carrot, cauliflower, celery, chives, corn (fresh), cucumber, eggplant, garlic stalk, hispid yam leaf, hyacinth bean (green), hot pepper, kidney bean (green), lettuce, lily, mung bean sprounts, mustard root, pumpkin, pumpkin sprouts, radish, radish leaf, snap pea, soybean sprouts, spinach, sweet pepper, tomato, turnip, water spinach, winter melon, yardlong cowpea
Mushrooms/seaweed	Button mushroom, “gold needle” mushroom, kelp, laver, oyster mushroom, shitake mushroom, silver ear fungus, wood ear fungus
Fruits	Apple, banana, casaba, date, dragon fruit, durian, gooseberry, grape, jujube, kiwi fruit, longan, orange, pear, persimmon, pitaya, pomegranate, pomelo, tangerine, watermelon
Nuts/seeds	Chestnut, lotus seed, peanut, pumpkin seed, sesame (black, white), sunflower seed, walnut, watermelon seed

Table S6.2. Gut microbiota that are found to have abilities to produce short-chain fatty acids (SCFAs) in previous studies and identified in the metagenome analysis sample

Phylum	Family	Species	Acetate[1]	Butyrate[1]	Propionate[1]	References
Present in at least 25% of the sample						
Actinobacteria	Bifidobacteriaceae	*Bifidobacterium longum*	Y	Y	Y	[81,184]
Bacteroidetes	Bacteroidaceae	*Bacteroides uniformis*			Y	[185]
		Bacteroides vulgatus			Y	[185,186]
	Prevotellaceae	*Prevotella copri*			Y	[185]
	Rikenellaceae	*Alistipes putredinis*		Y	Y	[185,187]
Firmicutes	Erysipelotrichaceae	*Clostridium innocuum*		Y		[159]
		Eubacterium biforme		Y		[159,185,187]
	Eubacteriaceae	*Eubacterium hallii*		Y	Y	[159,185-191]
		Eubacterium ramulus		Y		[159]
		Eubacterium rectale		Y		[184-188,190]
		Eubacterium ventriosum		Y		[159,184]
	Lachnospiraceae	*Anaerostipes hadrus*		Y		[185,186]
		Coprococcus catus		Y	Y	[159,185-187]
		Coprococcus comes		Y		[186,187]

		Lachnospiraceae bacterium 5_1_63FAA		Y		[187]
		Roseburia hominis	Y	Y		[192]
		Roseburia intestinalis		Y	Y	[159,185-187,190,193]
		Roseburia inulinivorans		Y		[185-188]
		Ruminococcus gnavus			Y	[186,194]
		Ruminococcus obeum			Y	[186]
		Ruminococcus torques			Y	[186]
	Lactobacillaceae	*Lactobacillus gasseri*			Y	[81]
	Ruminococcaceae	*Faecalibacterium prausnitzii*		Y		[159,184-186,188,190,193,195]
		Ruminococcus bromii	Y			[196]
	Veillonellaceae	*Dialister invisus*			Y	[185]
		Megamonas funiformis	Y		Y	[197]
		Veillonella parvula			Y	[186]
		Present in less than 25% of the sample				
Actinobacteria	Bifidobacteriaceae	*Bifidobacterium adolescentis*	Y			[191]
		Bifidobacterium bifidum	Y		Y	[81]

Bacteroidetes	Bacteroidaceae	*Bacteroides thetaiotaomicron*	Y		Y	[186,195]
	Porphyromonoadaceae	*Odoribacter splanchnicus*		Y		[187]
		Porphyromonas gingivalis		Y		[187]
Firmicutes	Clostridiaceae	*Clostridium beijerinckii*		Y		[159]
		Clostridium butyricum		Y		[159]
		Clostridium sp. L2-50		Y		[186]
		Clostridium symbiosum		Y		[159,187]
	Eubacteriaceae	*Anaerofustis stercorihominis*		Y		[187]
		Eubacterium limosum	Y	Y		[159,198,199]
	Erysipelotrichaceae	*Eubacterium cylindroides*		Y		[159]
		Eubacterium dolichum		Y		[159]
	Lachnospiraceae	*Anaerostipes caccae*		Y		[159,187,188,190,191,200]
		Blautia hydrogenotrophica	Y			[196]
		Butyrivibrio crossotus		Y		[159,187]
		Coprococcus eutactus		Y		[159,186,187]
	Lactobacillaceae	*Lactobacillus acidophilus*	Y	Y	Y	[201]
		Lactobacillus rhamnosus			Y	[81]
	Peptostreptococcaceae	*Clostridium difficile*		Y		[159,187]

	Ruminococcaceae	*Anaerotruncus colihominis*		Y		[187]
		Ruminococcaceae bacterium D16		Y		[187]
		Subdoligranulum variabile		Y		[185,187]
	Veillonellacea	*Megasphaera elsdenii*			Y	[186]
Fusobacteria	Fusobacteriaceae	*Fusobacterium mortiferum*		Y		[187]
		Fusobacterium nucleatum		Y		[187]
		Fusobacterium ulcerans		Y		[187]
		Fusobacterium varium		Y		[187]
Verrucomicrobia	Verrucomicrobiaceae	*Akkermansia muciniphila*	Y		Y	[185,202]

[1] "Y" indicates that the microbiota produces the given SCFA and empty cell indicates that there is no evidence showing that the microbiota produces the given SCFA.

Table S6.3. The interaction between dietary precursors of short-chain fatty acids (SCFAs) and overweight in models of plasma SCFAs [% change (95% CI)], corresponding to Figure 6.1[1]

n=490	Butyrate/isobutyrate	Valerate	Isovalerate	Total SCFAs
		Insoluble fiber (ref. = low)[2]		
Middle	-15.45 (-28.4, -0.15)*	-4.23 (-35.24, 41.63)	-18.02 (-33.48, 1.05)	-14.91 (-28.17, 0.8)
High	1.47 (-14.54, 20.47)	0.87 (-32.66, 51.09)	1.47 (-18.24, 25.92)	4.52 (-12.25, 24.51)
Middle X High BMI	25.71 (0.29, 57.55)*	7.83 (-36.62, 83.46)	31.45 (-1.05, 74.62)	24.92 (-0.75, 57.25)
High X High BMI	2.87 (-17.84, 28.79)	19.41 (-29.64, 102.64)	6.38 (-19.81, 41.12)	6.04 (-15.67, 33.32)
Interaction p-value[3]	0.095	0.803	0.140	0.144
		Carbohydrate (ref. = low)[2]		
Middle	4.89 (-11.35, 24.1)	22.43 (-17.39, 81.44)	4.17 (-15.66, 28.66)	5.37 (-11.17, 24.98)
High	-4.37 (-20.71, 15.34)	-8.9 (-41.23, 41.21)	-10.78 (-29.48, 12.89)	-8.44 (-24.3, 10.75)
Middle X High BMI	0.08 (-20.31, 25.7)	-30.31 (-59.11, 18.77)	0.91 (-24.2, 34.34)	-2.28 (-22.46, 23.16)
High X High BMI	12.38 (-10.4, 40.95)	12.18 (-33.96, 90.54)	19.29 (-10.24, 58.52)	21.55 (-3.42, 52.97)
Interaction p-value[3]	0.508	0.188	0.391	0.124
		High-fiber foods (ref. = low)[2]		
Middle	-7.4 (-21.45, 9.17)	4.64 (-28.92, 54.04)	-8.02 (-25.17, 13.06)	-7.73 (-21.95, 9.09)
High	-5.04 (-20.02, 12.74)	21.03 (-19.14, 81.13)	-4.25 (-22.79, 18.73)	-2.78 (-18.35, 15.76)

Middle X High BMI	-7.72 (-26.31, 15.55)	-10.56 (-47.27, 51.7)	-14.76 (-35.7, 13)	-9.54 (-28.04, 13.7)
High X High BMI	-1.63 (-21.47, 23.22)	-27.02 (-57.01, 23.88)	-1.46 (-25.7, 30.69)	-5.39 (-24.76, 18.97)
Interaction p-value[3]	0.761	0.497	0.472	0.689

[1] High BMI, BMI$\geqslant$24 kg/m^2. The abundance of each plasma SCFAs and the sum of the all three SCFAs were $\log_2$ transformed. The % change relative to the reference level was calculated using the following formula: (2^(β)-1)*100%, where β was the linear model coefficient. Model was adjusted for age, sex, batch, province, urbanization, income, education, total energy, insoluble fiber intake, physical activity, smoking, and alcohol intake. *, % change p-value<0.05, **, % change p-value<0.01

[2] Dietary intakes were measured by 3-consecutive 24h dietary recalls and household food inventories. High-fiber foods was calculated as the sum of whole grains, legumes, starchy roots, vegetables, mushrooms/seaweeds, fruits, nuts/seeds. Insoluble fiber, carbohydrate, and high-fiber food score were categorized by tertiles to represent low, middle, and high intakes.

[3] The statistical significance of the interaction term between dietary factors and overweight was estimated using a Wald test that compared models with and without the interaction term. .

Table S6.4. The interaction between dietary precursors of short-chain fatty acids (SCFAs) and abdominal obesity in models of plasma SCFAs [% change (95% CI)], corresponding to Figure 6.2[1]

n=488	Butyrate/isobutyrate	Valerate	Isovalerate	Total SCFAs
		Insoluble fiber (ref. = low)[2]		
Middle	-18.42 (-33.00, -0.66)*	-9.96 (-43.57, 43.66)	-19.07 (-36.9, 3.8)	-18.25 (-33.17, -0.01)*
High	6.31 (-12.88, 29.73)	1.72 (-36.57, 63.13)	9.63 (-14.76, 41)	9.02 (-11.06, 33.64)
Middle X High WHtR	23.88 (-2.03, 56.63)	14.72 (-34.24, 100.13)	23.52 (-8.17, 66.14)	23.79 (-2.62, 57.35)
High X High WHtR	-5.15 (-24.96, 19.9)	8.45 (-37.8, 89.1)	-7.08 (-30.9, 24.95)	-2.87 (-23.57, 23.44)
Interaction p-value[3]	0.063	0.888	0.150	0.098
		Carbohydrate (ref. = low)[2]		
Middle	3.24 (-15.5, 26.14)	-17.26 (-48.32, 32.47)	-0.02 (-22.34, 28.71)	-5.73 (-23.18, 15.7)
High	9.64 (-11.83, 36.35)	-41.01 (-64.65, -1.56)*	5.98 (-19.49, 39.51)	-4.46 (-23.54, 19.38)
Middle X High WHtR	2.91 (-18.84, 30.49)	36.07 (-22.1, 137.67)	7.41 (-20.38, 44.9)	16.27 (-8.78, 48.21)
High X High WHtR	-11.21 (-30.08, 12.75)	100.19 (14.22, 250.87)*	-12.13 (-34.99, 18.75)	7.57 (-15.73, 37.32)
Interaction p-value[3]	0.428	0.052	0.404	0.475
		High-fiber foods (ref. = low)[2]		
Middle	-14.88 (-29.94, 3.41)	10.78 (-30.14, 75.66)	-17.55 (-35.48, 5.37)	-12.92 (-28.62, 6.23)
High	-3.71 (-21.12, 17.54)	8.02 (-32.65, 73.24)	-0.95 (-22.97, 27.34)	-2.98 (-20.87, 18.94)

Middle X High WHtR	6.66 (-15.49, 34.62)	-14.69 (-50.85, 48.09)	4.3 (-22.22, 39.86)	1.14 (-20.27, 28.29)
High X High WHtR	-4.84 (-24.57, 20.03)	-2.69 (-43.87, 68.69)	-7.95 (-31.31, 23.35)	-4.77 (-24.89, 20.73)
Interaction p-value[3]	0.634	0.835	0.702	0.872

[1] Waist-to-height ratio $\geqslant$0.5. The abundance of each SCFAs and the sum of the all three SCFAs were $\log_2$ transformed. The % change relative to the reference level was calculated using the following formula: (2^(β)-1)*100%, where β was the linear model coefficient. Model was adjusted for age, sex, batch, province, urbanization, income, education, total energy, insoluble fiber, physical activity, smoking, and alcohol intake. *, % change p-value<0.05, **, % change p-value<0.01.

[2] Dietary intakes were measured by 3-consecutive 24h dietary recalls and household food inventories. High-fiber foods was calculated as the sum of whole grains, legumes, starchy roots, vegetables, mushrooms/seaweeds, fruits, nuts/seeds. Insoluble fiber, carbohydrate, and high-fiber food score were categorized by tertiles to represent low, middle, and high intakes.

[3] The statistical significance of the interaction term between dietary factors and abdominal obesity was estimated using a Wald test that compared models with and without the interaction term

Table S6.5. The associations between individual high-fiber foods and plasma short-chain fatty acids (SCFAs) by overweight (n=490)[1]

	Butyrate/isobutyrate		Valerate		Isovalerate		Total SCFAs	
	Interaction p-value[2]	**% change (95% CI)**	**Interaction p-value[2]**	**% change (95% CI)**	**Interaction p-value[2]**	**% change (95% CI)**	**Interaction p-value[2]**	**% change (95% CI)**
Wholegrains (ref. = non-consumer)[3]								
Normal BMI	0.64	-9.24 (-30.92, 19.26)	0.36	25.87 (-33.52, 138.32)	0.79	-11.06 (-36.91, 25.39)	0.3	-4.78 (-27.92, 25.79)
High BMI		-17.71 (-39.27, 11.51)		-18.75 (-60.07, 65.35)		-16.92 (-43.31, 21.76)		-23.38 (-43.79, 4.45)
Legumes (ref. = below median)[3]								
Normal BMI	0.56	7.66 (-5.8, 23.04)	0.06	21.84 (-10.74, 66.31)	0.42	9.61 (-7.34, 29.65)	0.21	9.49 (-4.45, 25.47)
High BMI		1.82 (-11.17, 16.73)		-19.54 (-41.47, 10.59)		-0.45 (-16.15, 18.2)		-2.92 (-15.54, 11.59)
Starchy roots (ref. = below median)[3]								
Normal BMI	0.89	-5.47 (-17.1, 7.79)	0.8	-5.75 (-30.65, 28.11)	0.74	-3.33 (-18.05, 14.02)	0.79	-7.11 (-18.74, 6.19)
High BMI		-6.67		-10.88		-7		-9.41

		(-18.45, 6.81)		(-34.98, 22.17)		(-21.52, 10.21)		(-21.05, 3.94)
Vegetables (ref. = below median)[3]								
Normal BMI	0.44	-3.44	0.51	7.73	0.32	-4.62	0.51	-2.93
		(-15.31, 10.1)		(-20.7, 46.35)		(-19.12, 12.47)		(-15.1, 10.98)
High BMI		3.95		-6.94		7.46		3.46
		(-9.21, 19)		(-32.15, 27.65)		(-9.34, 27.38)		(-9.89, 18.78)
Mushrooms/seaweeds (ref. = non-consumer)[3]								
Normal BMI	0.32	3.86	0.32	-16.28	0.47	7.97	0.65	1.85
		(-9.6, 19.31)		(-39.44, 15.74)		(-9.31, 28.54)		(-11.61, 17.35)
High BMI		-6.01		5.94		-1.53		-2.73
		(-18.84, 8.86)		(-24.8, 49.26)		(-18.13, 18.43)		(-16.28, 13)
Fruits (ref. = non-consumer)[3]								
Normal BMI	0.89	-6.54	0.04	4.8	0.9	-5.36	0.21	-2.2
		(-18.43, 7.08)		(-23.61, 43.79)		(-20.25, 12.31)		(-14.85, 12.34)
High BMI		-7.76		-32.81		-6.76		-13.15
		(-19.58, 5.79)		(-51.15, -7.58)*		(-21.54, 10.8)		(-24.47, -0.13)*
Nuts/seeds (ref. = non-consumer)[3]								
Normal BMI	0.69	11.37	0.78	-1.14	0.61	17.98	0.31	17.13

	(-5.64, 31.45)	(-32.92, 45.71)	(-4.19, 45.27)	(-1.07, 38.69)
High BMI	6.19	6.78	9.33	3.35
	(-10.59, 26.11)	(-28.6, 59.67)	(-11.9, 35.69)	(-13.26, 23.15)

[1] High BMI, BMI⩾24 kg/m^2. The abundance of each SCFAs and the sum of the all three SCFAs were $\log_2$ transformed. The % change relative to the reference level was calculated using the following formula: (2^(β)-1)*100%, where β was the linear model coefficient. Model was adjusted for age, sex, batch, province, urbanization, income, education, total energy, insoluble fiber, physical activity, smoking, and alcohol intake. *, % change p-value<0.05.

[2] The statistical significance of the interaction term between dietary factors and overweight was estimated using a Wald test that compared models with and without the interaction term.

[3] Dietary intakes were measured by 3-consecutive 24h dietary recalls and household food inventories. Foods consumed by more than and less than 50% of the analysis sample were categorized by median and yes/no consumers, respectively.

Table S6.6. The associations between individual high-fiber foods and plasma short-chain fatty acids (SCFAs) by abdominal obesity (n=488)[1]

	Butyrate/isobutyrate		**Valerate**		**Isovalerate**		**Total SCFAs**	
	Interaction p-value[2]	**% change (95% CI)**	**Interaction p-value[2]**	**% change (95% CI)**	**Interaction p-value[2]**	**% change (95% CI)**	**Interaction p-value[2]**	**% change (95% CI)**
Wholegrains (ref. = non-consumer)[3]								
Normal	0.098	9.32	0.005	164.02	0.16	10.65	0.003	31.64
WHTR		(-22.51, 54.23)		(17.88, 491.34)*		(-28.38, 70.95)		(-7.26, 86.87)
High		-23.35		-36.26		-24.58		-30.84
WHTR		(-40.14, -1.85)*		(-64.28, 13.74)		(-44.82, 3.07)		(-46.23, -11.06)**
Legumes (ref. = below median)[3]								
Normal	0.45	9.81	0.98	-0.07	0.42	11.39	0.51	7.57
WHTR		(-6.22, 28.57)		(-31.07, 44.89)		(-8.72, 35.94)		(-8.49, 26.45)
High		1.97		-0.53		0.83		0.78
WHTR		(-9.38, 14.73)		(-24.65, 31.33)		(-13.13, 17.02)		(-10.7, 13.74)
Starchy roots (ref. = below median)[3]								
Normal	0.95	-5.83	0.42	5.93	0.87	-3.99	0.71	-5.82
WHTR		(-19.74, 10.48)		(-27.27, 54.28)		(-21.54, 17.48)		(-20.02, 10.91)

High WHTR		-6.44 (-16.56, 4.91)		-12.1 (-32.86, 15.09)		-5.98 (-18.64, 8.66)		-9.25 (-19.29, 2.03)
Vegetables (ref. = below median)[3]								
Normal WHTR	0.95	-0.16 (-14.69, 16.84)	0.83	-4.23 (-33.82, 38.6)	0.82	3.13 (-15.43, 25.76)	0.95	0.48 (-14.46, 18.04)
High WHTR		0.44 (-10.51, 12.74)		0.67 (-23.26, 32.06)		0.33 (-13.27, 16.07)		-0.13 (-11.27, 12.41)
Mushrooms/seaweeds (ref. = non-consumer)[3]								
Normal WHTR	0.36	5.21 (-11.27, 24.76)	0.88	-8.7 (-38.85, 36.31)	0.34	12.2 (-9.5, 39.1)	0.52	4.07 (-12.61, 23.93)
High WHTR		-4.52 (-15.72, 8.17)		-5.13 (-29.26, 27.23)		-1.31 (-15.68, 15.52)		-2.9 (-14.55, 10.35)
Fruits (ref. = non-consumer)[3]								
Normal WHTR	0.28	-0.48 (-15.42, 17.11)	0.001	39.98 (-4.17, 104.47)	0.36	1.05 (-17.74, 24.13)	0.02	7.2 (-9.19, 26.56)

High WHTR		-10.49 (-20.55, 0.84)		-33.35 (-49.51, -12.03)**		-9.66 (-22.3, 5.04)		-14.37 (-24.18, -3.3)*
Nuts/seeds (ref. = non-consumer)[3]								
Normal WHTR	0.09	24.46 (1.69, 52.34)	0.91	6.65 (-33.84, 71.91)	0.02	43.57 (11.35, 85.13)	0.03	32.35 (7.66, 62.71)*
High WHTR		0.39 (-13.32, 16.26)		2.98 (-27.2, 45.65)		-0.61 (-17.36, 19.54)		-0.56 (-14.41, 15.52)

[1] Waist-to-height ratio ⩾0.5. The abundance of each SCFAs and the sum of the all three SCFAs were $\log_2$ transformed. The % change relative to the reference level was calculated using the following formula: (2^(β)-1)*100%, where β was the linear model coefficient. Model was adjusted for age, sex, batch, province, urbanization, income, education, total energy, insoluble fiber, physical activity, smoking, and alcohol intake. *, % change p-value<0.05; **, % change p-value<0.01.

[2] The statistical significance of the interaction term between dietary factors and abdominal obesity was estimated using a Wald test that compared models with and without the interaction term.

[3] Dietary intakes were measured by 3-consecutive 24h dietary recalls and household food inventories. Foods consumed by more than and less than 50% of the analysis sample were categorized by median and yes/no consumers, respectively.

Table S6.7. The associations between each 56 microbial short-chain fatty acid (SCFA) producers with body mass index (BMI) and waist-to-height ratio (WHtR)[1]

	BMI (n=209)		WHtR (n=208)		
Species	**β (95% confidence interval)**	**p-value**	**β (95% confidence interval)**	**p-value**	**Rare[2]**
Bifidobacterium longum	-0.03 (-0.24, 0.18)	0.78	3.E-06 (-4.E-03, 4.E-03)	1.00	N
Bacteroides thetaiotaomicron	-0.32 (-0.66, 0.01)	0.06	-7.E-03 (-0.01, -1.E-03)	0.02	N
Bacteroides uniformis	-0.07 (-0.4, 0.25)	0.66	-3.E-03 (-0.01, 3.E-03)	0.38	N
Bacteroides vulgatus	0.05 (-0.33, 0.42)	0.80	-3.E-04 (-0.01, 0.01)	0.92	N
Prevotella copri	0.08 (-0.16, 0.32)	0.51	3.E-03 (-2.E-03, 0.01)	0.25	N
Alistipes putredinis	0.15 (-0.14, 0.43)	0.31	3.E-04 (-5.E-03, 0.01)	0.92	N
Lactobacillus gasseri	-0.07 (-0.44, 0.31)	0.73	-1.E-04 (-0.01, 0.01)	0.97	N
Eubacterium hallii	0.54 (0.22, 0.87)	1.E-03	9.E-03 (3.E-03, 0.01)	0.01	N
Eubacterium ramulus	0.21 (-0.09, 0.51)	0.17	3.E-03 (-3.E-03, 0.01)	0.30	N
Eubacterium rectale	0.31 (0.08, 0.54)	0.01	6.E-03 (2.E-03, 0.01)	0.01	N
Eubacterium ventriosum	0.01 (-0.29, 0.31)	0.95	1.E-04 (-0.01, 0.01)	0.97	N
Anaerostipes hadrus	0.22 (-0.17, 0.61)	0.27	1.E-03 (-0.01, 0.01)	0.78	N
Ruminococcus gnavus	-0.04 (-0.29, 0.22)	0.78	3.E-05 (-5.E-03, 5.E-03)	0.99	N
Ruminococcus obeum	0.53 (-0.04, 1.11)	0.07	6.E-03 (-5.E-03, 0.02)	0.28	N

Ruminococcus torques	0.61 (-0.09, 1.31)	0.09	9.E-03 (-4.E-03, 0.02)	0.18	N
Coprococcus catus	0.03 (-0.25, 0.31)	0.83	2.E-03 (-3.E-03, 0.01)	0.44	N
Coprococcus comes	0.01 (-0.24, 0.27)	0.92	-2.E-05 (-5.E-03, 5.E-03)	0.99	N
Lachnospiraceae bacterium 5 1 63FAA	0.2 (-0.18, 0.58)	0.29	7.E-04 (-0.01, 0.01)	0.85	N
Roseburia hominis	0.17 (-0.13, 0.46)	0.27	2.E-04 (-0.01, 0.01)	0.95	N
Roseburia intestinalis	0.23 (-0.06, 0.51)	0.12	3.E-04 (-5.E-03, 0.01)	0.90	N
Roseburia inulinivorans	0.11 (-0.19, 0.41)	0.46	1.E-03 (-5.E-03, 0.01)	0.72	N
Faecalibacterium prausnitzii	0.17 (-0.23, 0.57)	0.40	4.E-03 (-4.E-03, 0.01)	0.34	N
Ruminococcus bromii	-0.13 (-0.34, 0.08)	0.23	-2.E-03 (-0.01, 2.E-03)	0.39	N
Clostridium innocuum	-0.05 (-0.57, 0.47)	0.85	2.E-04 (-0.01, 0.01)	0.97	N
Eubacterium biforme	0.08 (-0.14, 0.31)	0.47	3.E-03 (-1.E-03, 0.01)	0.15	N
Dialister invisus	-0.16 (-0.48, 0.17)	0.34	-3.E-03 (-0.01, 3.E-03)	0.37	N
Megamonas funiformis	0.45 (0.14, 0.76)	0.01	6.E-03 (-8.40E-05, 0.01)	0.05	N
Veillonella parvula	-0.16 (-0.55, 0.23)	0.42	-2.E-03 (-0.01, 5.E-03)	0.53	N
Akkermansia muciniphila	0.06 (-0.19, 0.31)	0.66	1.E-03 (-4.E-03, 0.01)	0.69	N
Bifidobacterium adolescentis	-0.09 (-1.33, 1.15)	0.88	7.E-03 (-0.02, 0.03)	0.56	Y
Bifidobacterium bifidum	-0.89 (-2.46, 0.68)	0.26	-1.E-02 (-0.04, 0.02)	0.45	Y
Odoribacter splanchnicus	-0.26 (-1.49, 0.98)	0.68	3.E-03 (-0.02, 0.03)	0.83	Y

Porphyromonas gingivalis	-1.5 (-4.44, 1.45)	0.32	2.E-02 (-0.03, 0.08)	0.43	Y
Lactobacillus acidophilus	0.46 (-4.14, 5.06)	0.84	2.E-02 (-0.07, 0.1)	0.68	Y
Lactobacillus rhamnosus	0.01 (-1.37, 1.4)	0.98	-8.E-03 (-0.03, 0.02)	0.54	Y
Clostridium beijerinckii	0.22 (-6.5, 6.94)	0.95	-7.E-02 (-0.19, 0.06)	0.29	Y
Clostridium butyricum	-1.14 (-3.86, 1.57)	0.41	-4.E-02 (-0.09, 0.01)	0.15	Y
Clostridium sp. L2 50	0.39 (-1.85, 2.62)	0.73	2.E-02 (-0.02, 0.06)	0.37	Y
Clostridium symbiosum	-2.08 (-3.22, -0.94)	4.E-04	-2.E-02 (-0.05, -4.E-03)	0.02	Y
Anaerofustis stercorihominis	0.9 (-0.67, 2.48)	0.26	1.E-02 (-0.02, 0.04)	0.41	Y
Eubacterium limosum	-0.01 (-1.25, 1.23)	0.99	-9.E-03 (-0.03, 0.01)	0.41	Y
Anaerostipes caccae	-0.94 (-3.02, 1.15)	0.38	-1.E-02 (-0.05, 0.03)	0.54	Y
Blautia hydrogenotrophica	-1.61 (-3.62, 0.4)	0.12	-4.E-02 (-0.07, -2.E-04)	0.05	Y
Butyrivibrio crossotus	-0.94 (-2.82, 0.95)	0.33	-7.E-03 (-0.04, 0.03)	0.68	Y
Coprococcus eutactus	0.5 (-0.99, 1.98)	0.51	1.E-02 (-0.02, 0.04)	0.44	Y
Clostridium difficile	0.07 (-1.62, 1.77)	0.93	6.E-03 (-0.02, 0.04)	0.70	Y
Anaerotruncus colihominis	-0.93 (-2.63, 0.77)	0.28	-2.E-02 (-0.05, 0.01)	0.19	Y
Ruminococcaceae bacterium D16	-3.36 (-6.61, -0.11)	0.04	-5.E-02 (-0.11, 0.01)	0.14	Y
Subdoligranulum variabile	-0.88 (-2.69, 0.92)	0.34	-2.E-02 (-0.05, 0.02)	0.37	Y
Eubacterium cylindroides	-0.77 (-2.81, 1.28)	0.46	-4.E-03 (-0.04, 0.03)	0.83	Y

Eubacterium dolichum	0.71 (-0.65, 2.07)	0.31	3.E-03 (-0.02, 0.03)	0.79	Y
Megasphaera elsdenii	-0.41 (-4.95, 4.14)	0.86	4.E-02 (-0.04, 0.12)	0.36	Y
Fusobacterium mortiferum	-0.06 (-1.21, 1.09)	0.92	5.E-04 (-0.02, 0.02)	0.96	Y
Fusobacterium nucleatum	-0.55 (-2.1, 0.99)	0.48	2.E-02 (-0.01, 0.04)	0.27	Y
Fusobacterium ulcerans	-0.62 (-3.59, 2.35)	0.68	-9.E-03 (-0.06, 0.05)	0.75	Y
Fusobacterium varium	-4.78 (-7.92, -1.64)	3.E-03	-6.E-02 (-0.12, -4.E-03)	0.04	Y

[1] The 56 microbial SCFA producers were selected from literature and the full list with references is in Table S2. The raw counts of each species were normalized and $\log_{10}$ transformed. Model was adjusted for age, sex, batch, province, urbanization, income, education, total energy, insoluble fiber, physical activity, smoking, and alcohol intake. β coefficients are interpreted as units of BMI or WHtR associated with per 1 unit increase in log10 of the relative abundance of microbiota.

[2] Microbiota presented in at least 25% of the sample were denoted as non-rare microbiota and kept as continuous variables in linear regression. Microbiota present in less than 25% of the sample were denoted as rare microbiota and included as binary variables (yes/no present in the sample) in linear regression.

Table S6.8. The associations between the overall and the total relative abundance of 56 microbial short-chain fatty acid (SCFA) producers with plasma SCFAs (n=209).

	Overall[1]		Total[2]	
	R^2	**P-value**	**β (95% CI)**	**P-value**
Butyrate/isobutyrate	0.002	0.95	-0.01 (-0.39, 0.37)	0.97
Valerate	0.004	0.58	0.54 (-0.37, 1.46)	0.24
Isovalerate	0.003	0.79	0.00 (-0.49, 0.48)	0.99
Total SCFAs	0.003	0.82	0.06 (-0.33, 0.45)	0.77

The 56 microbial SCFA producers were selected from literature and the full list with references is in Table S2. The relative abundance of each SCFAs and the sum of the all three SCFAs were $\log_2$ transformed. The raw counts of each species and the total counts of the 56 species were normalized and $\log_{10}$ transformed. Model was adjusted for age, sex, batch, province, urbanization, income, education, total energy, insoluble fiber, physical activity, smoking, and alcohol intake.

[1] R^2 and p-value were calculated using permutational multivariate analysis of variance (PERMANOVA) of all 56 species.

[2] Linear regression was performed on the total relative abundance of the 56 species.

Table S6.9. The associations between each 56 microbial short-chain fatty acid (SCFA) producers with plasma SCFAs (n=209)[1]

	Butyrate/isovalerate		Valerate		Isovalerate		Total SCFAs		
Species	**β (95% confidence interval)**	**p**	**β (95% confidence interval)**	**p**	**β (95% confidence interval)**	**p**	**β (95% confidence interval)**	**p**	**Rare[2]**
Bifidobacterium longum	0 (-0.05, 0.04)	0.84	-0.04 (-0.16, 0.07)	0.46	-0.02 (-0.08, 0.04)	0.54	-0.02 (-0.07, 0.03)	0.45	N
Bacteroides thetaiotaomicron	-0.04 (-0.12, 0.04)	0.29	0.02 (-0.16, 0.21)	0.81	-0.04 (-0.14, 0.05)	0.38	-0.04 (-0.12, 0.04)	0.36	N
Bacteroides uniformis	0 (-0.07, 0.08)	0.92	-0.13 (-0.3, 0.05)	0.17	0 (-0.09, 0.1)	0.92	-0.03 (-0.1, 0.05)	0.50	N
Bacteroides vulgatus	-0.03 (-0.12, 0.05)	0.47	-0.08 (-0.28, 0.13)	0.46	-0.04 (-0.14, 0.07)	0.52	-0.04 (-0.13, 0.04)	0.33	N
Prevotella copri	-0.03 (-0.08, 0.03)	0.31	0 (-0.13, 0.13)	0.99	-0.03 (-0.1, 0.04)	0.42	-0.02 (-0.08, 0.03)	0.41	N
Alistipes putredinis	-0.02 (-0.08, 0.04)	0.54	0.04 (-0.11, 0.2)	0.59	-0.03 (-0.11, 0.05)	0.48	-0.02 (-0.08, 0.05)	0.64	N
Lactobacillus gasseri	-0.08 (-0.16, 0.01)	0.07	-0.03 (-0.24, 0.18)	0.78	-0.09 (-0.19, 0.02)	0.12	-0.09 (-0.18, 0)	0.04	N
Eubacterium hallii	0 (-0.08, 0.07)	0.93	0 (-0.18, 0.18)	0.99	-0.02 (-0.12, 0.07)	0.65	-0.02 (-0.09, 0.06)	0.67	N
Eubacterium ramulus	-0.05 (-0.11, 0.02)	0.19	0.03 (-0.14, 0.2)	0.73	-0.06 (-0.15, 0.02)	0.16	-0.04 (-0.11, 0.03)	0.31	N
Eubacterium rectale	0 (-0.05, 0.05)	0.99	0.09 (-0.04, 0.22)	0.16	0 (-0.07, 0.06)	0.93	0 (-0.05, 0.06)	0.94	N

Eubacterium ventriosum	-0.05 (-0.12, 0.02)	0.18	0.01 (-0.16, 0.18)	0.91	-0.06 (-0.15, 0.03)	0.16	-0.05 (-0.12, 0.02)	0.18	N
Anaerostipes hadrus	0 (-0.09, 0.09)	0.97	0.08 (-0.14, 0.3)	0.46	-0.01 (-0.12, 0.11)	0.89	0.02 (-0.07, 0.11)	0.71	N
Ruminococcus gnavus	0.05 (0, 0.11)	0.07	0.05 (-0.09, 0.18)	0.52	0.07 (0, 0.14)	0.05	0.07 (0.01, 0.13)	0.03	N
Ruminococcus obeum	-0.11 (-0.25, 0.02)	0.10	0.15 (-0.17, 0.48)	0.35	-0.16 (-0.33, 0.01)	0.06	-0.07 (-0.21, 0.07)	0.33	N
Ruminococcus torques	0.03 (-0.13, 0.19)	0.74	0.27 (-0.12, 0.65)	0.17	0.05 (-0.16, 0.25)	0.66	0.1 (-0.06, 0.26)	0.24	N
Coprococcus catus	-0.03 (-0.09, 0.04)	0.39	-0.04 (-0.2, 0.11)	0.58	-0.04 (-0.12, 0.04)	0.28	-0.04 (-0.1, 0.03)	0.26	N
Coprococcus comes	0.02 (-0.04, 0.08)	0.55	-0.01 (-0.15, 0.13)	0.93	0.02 (-0.05, 0.09)	0.60	0.02 (-0.04, 0.07)	0.60	N
Lachnospiraceae bacterium 5 1 63FAA	0.06 (-0.03, 0.15)	0.18	0.1 (-0.11, 0.31)	0.36	0.08 (-0.03, 0.19)	0.16	0.08 (-0.01, 0.17)	0.07	N
Roseburia hominis	-0.05 (-0.11, 0.02)	0.17	0.07 (-0.1, 0.23)	0.42	-0.06 (-0.15, 0.02)	0.15	-0.04 (-0.11, 0.02)	0.21	N
Roseburia intestinalis	-0.02 (-0.08, 0.05)	0.62	-0.19 (-0.35, -0.04)	0.01	-0.02 (-0.1, 0.06)	0.59	-0.05 (-0.12, 0.01)	0.11	N
Roseburia inulinivorans	0.01 (-0.06, 0.08)	0.78	-0.11 (-0.27, 0.06)	0.21	0.01 (-0.08, 0.1)	0.83	-0.02 (-0.09, 0.05)	0.64	N
Faecalibacterium prausnitzii	-0.09 (-0.18, 0)	0.06	-0.06 (-0.28, 0.16)	0.61	-0.14 (-0.25, -0.02)	0.02	-0.11 (-0.21, -0.02)	0.02	N
Ruminococcus bromii	-0.03 (-0.07, 0.02)	0.30	-0.02 (-0.14, 0.1)	0.75	-0.04 (-0.1, 0.02)	0.20	-0.03 (-0.08, 0.02)	0.24	N

Clostridium innocuum	0.02 (-0.1, 0.14)	0.70	0.25 (-0.03, 0.54)	0.08	0.04 (-0.11, 0.19)	0.64	0.06 (-0.06, 0.18)	0.33	N
Eubacterium biforme	0 (-0.05, 0.06)	0.88	0.09 (-0.03, 0.22)	0.16	0 (-0.06, 0.07)	0.89	0.02 (-0.04, 0.07)	0.51	N
Dialister invisus	-0.04 (-0.12, 0.03)	0.25	0.11 (-0.07, 0.28)	0.25	-0.06 (-0.16, 0.03)	0.19	-0.02 (-0.1, 0.05)	0.57	N
Megamonas funiformis	-0.03 (-0.1, 0.04)	0.39	0.01 (-0.16, 0.19)	0.89	-0.02 (-0.11, 0.07)	0.63	-0.02 (-0.09, 0.06)	0.62	N
Veillonella parvula	0.06 (-0.03, 0.15)	0.19	-0.03 (-0.25, 0.19)	0.80	0.1 (-0.02, 0.21)	0.09	0.05 (-0.04, 0.14)	0.31	N
Akkermansia muciniphila	-0.03 (-0.09, 0.02)	0.23	0.07 (-0.06, 0.21)	0.29	-0.05 (-0.12, 0.02)	0.19	-0.02 (-0.08, 0.03)	0.41	N
Bifidobacterium adolescentis	0.11 (-0.18, 0.39)	0.46	-0.12 (-0.82, 0.57)	0.73	0.14 (-0.22, 0.51)	0.45	0.08 (-0.21, 0.38)	0.59	Y
Bifidobacterium bifidum	0.01 (-0.35, 0.37)	0.96	-0.54 (-1.4, 0.33)	0.22	0.04 (-0.41, 0.5)	0.85	-0.03 (-0.4, 0.34)	0.88	Y
Odoribacter splanchnicus	-0.12 (-0.4, 0.16)	0.39	-0.05 (-0.73, 0.63)	0.89	-0.17 (-0.53, 0.19)	0.34	-0.15 (-0.44, 0.13)	0.29	Y
Porphyromonas gingivalis	0.36 (-0.32, 1.03)	0.30	0.57 (-1.07, 2.21)	0.49	0.43 (-0.43, 1.29)	0.33	0.46 (-0.23, 1.15)	0.19	Y
Lactobacillus acidophilus	0.5 (-0.56, 1.55)	0.35	2.08 (-0.46, 4.62)	0.11	0.49 (-0.85, 1.83)	0.47	0.61 (-0.47, 1.69)	0.27	Y

Lactobacillus rhamnosus	-0.17 (-0.48, 0.15)	0.29	0.34 (-0.43, 1.1)	0.39	-0.25 (-0.65, 0.15)	0.22	-0.19 (-0.51, 0.14)	0.25	Y
Clostridium beijerinckii	-0.32 (-1.85, 1.2)	0.68	1.94 (-1.75, 5.62)	0.30	-0.15 (-2.09, 1.79)	0.88	0.18 (-1.39, 1.75)	0.83	Y
Clostridium butyricum	0.36 (-0.26, 0.97)	0.26	-0.82 (-2.31, 0.68)	0.28	0.53 (-0.25, 1.31)	0.18	0.28 (-0.36, 0.91)	0.39	Y
Clostridium sp. L2 50	0.34 (-0.18, 0.86)	0.20	1.04 (-0.21, 2.3)	0.10	0.41 (-0.25, 1.07)	0.23	0.48 (-0.05, 1.01)	0.08	Y
Clostridium symbiosum	-0.12 (-0.39, 0.15)	0.37	0.58 (-0.07, 1.22)	0.08	-0.07 (-0.41, 0.27)	0.69	0 (-0.27, 0.28)	0.99	Y
Anaerofustis stercorihominis	0.08 (-0.27, 0.44)	0.64	-0.72 (-1.58, 0.14)	0.10	0.02 (-0.44, 0.47)	0.95	-0.12 (-0.48, 0.25)	0.54	Y
Eubacterium limosum	0.04 (-0.25, 0.32)	0.80	-0.24 (-0.92, 0.44)	0.49	0.13 (-0.23, 0.48)	0.48	0.07 (-0.22, 0.36)	0.65	Y
Anaerostipes caccae	-0.14 (-0.61, 0.33)	0.56	-0.13 (-1.27, 1.02)	0.83	-0.2 (-0.8, 0.4)	0.51	-0.19 (-0.68, 0.3)	0.44	Y
Blautia hydrogenotrophica	-0.06 (-0.52, 0.4)	0.80	-0.29 (-1.4, 0.81)	0.60	-0.09 (-0.67, 0.49)	0.75	-0.08 (-0.55, 0.39)	0.73	Y
Butyrivibrio crossotus	-0.25 (-0.68, 0.17)	0.24	0.67 (-0.36, 1.7)	0.20	-0.41 (-0.95, 0.13)	0.13	-0.2 (-0.63, 0.24)	0.38	Y
Coprococcus eutactus	0.2 (-0.13, 0.54)	0.23	0.38 (-0.44, 1.2)	0.36	0.26 (-0.17, 0.69)	0.24	0.26 (-0.09, 0.61)	0.14	Y
Clostridium difficile	0.17 (-0.22, 0.56)	0.39	0.15 (-0.79, 1.09)	0.75	0.32 (-0.17, 0.81)	0.20	0.22 (-0.18, 0.61)	0.28	Y

Anaerotruncus colihominis	-0.31 (-0.7, 0.07)	0.11	-0.58 (-1.51, 0.35)	0.22	-0.31 (-0.79, 0.18)	0.22	-0.37 (-0.76, 0.02)	0.07	Y
Ruminococcaceae bacterium D16	0.24 (-0.51, 0.99)	0.53	1.46 (-0.35, 3.27)	0.11	0.29 (-0.67, 1.24)	0.55	0.6 (-0.17, 1.37)	0.13	Y
Subdoligranulum variabile	0 (-0.41, 0.41)	1.00	1.18 (0.19, 2.17)	0.02	-0.09 (-0.62, 0.43)	0.72	0.16 (-0.26, 0.59)	0.45	Y
Eubacterium cylindroides	-0.13 (-0.6, 0.34)	0.58	0.45 (-0.69, 1.59)	0.44	-0.26 (-0.85, 0.34)	0.40	-0.06 (-0.54, 0.42)	0.81	Y
Eubacterium dolichum	0.32 (0, 0.64)	0.05	-0.33 (-1.1, 0.45)	0.41	0.43 (0.02, 0.83)	0.04	0.24 (-0.09, 0.57)	0.15	Y
Megasphaera elsdenii	0.43 (-0.59, 1.46)	0.41	1.98 (-0.49, 4.45)	0.12	0.51 (-0.79, 1.81)	0.44	0.67 (-0.38, 1.72)	0.21	Y
Fusobacterium mortiferum	0.18 (-0.08, 0.43)	0.18	0.05 (-0.58, 0.68)	0.87	0.28 (-0.05, 0.61)	0.10	0.27 (0.01, 0.54)	0.04	Y
Fusobacterium nucleatum	0.02 (-0.33, 0.37)	0.93	0.26 (-0.58, 1.11)	0.54	-0.04 (-0.48, 0.4)	0.86	0.06 (-0.3, 0.41)	0.76	Y
Fusobacterium ulcerans	0.59 (-0.07, 1.26)	0.08	1.08 (-0.53, 2.7)	0.19	0.8 (-0.05, 1.64)	0.07	0.76 (0.08, 1.44)	0.03	Y
Fusobacterium varium	0.28 (-0.46, 1.01)	0.46	1.02 (-0.76, 2.8)	0.26	0.5 (-0.44, 1.43)	0.30	0.51 (-0.24, 1.27)	0.18	Y

[1] The 56 microbial SCFA producers were selected from literature and the full list with references is in Table S2. The relative abundance of each SCFAs and the sum of the all three SCFAs were $\log_2$ transformed. The raw counts of each species and the total counts of the 56 species were

normalized and $\log_{10}$ transformed. Model was adjusted for age, sex, batch, province, urbanization, income, education, total energy, insoluble fiber, physical activity, smoking, and alcohol intake. β coefficients can be used in the following formula: (2^(β)-1)*100% to estimate the percent change in butyrate associated with per 1 unit increase in $\log_{10}$ of the total relative abundance of SCFAs-producing gut microbiota.

[2] Microbiota presented in at least 25% of the sample were denoted as non-rare microbiota and kept as continuous variables in linear regression. Microbiota present in less than 25% of the sample were denoted as rare microbiota and included as binary variables (yes/no present in the sample) in linear regression.

Table S6.10. Sensitivity analysis of the associations between plasma short-chain fatty acids (SCFAs) with body mass index (BMI) and waist-to-height ratio (WHtR)

		BMI (n=462)		WHtR (n=460)	
	Mean (SD)	**β (95% confidence interval)**	**p-value**	**β (95% confidence interval)**	**p-value**
Butyrate/isobutyrate	-0.03 (0.76)	0.36 (-0.02, 0.75)	0.07	0.01 (3E-03, 0.02)	0.01
Valerate	-0.11 (1.74)	0.01 (-0.16, 0.18)	0.91	2E-03 (-2E-03, 4E-03)	0.36
Isovalerate	0.08 (0.94)	0.18 (-0.13, 0.50)	0.24	0.01 (-4E-04, 0.01)	0.07
Total SCFAs	1.51 (0.76)	0.27 (-0.11, 0.65)	0.17	0.01 (2E-03, 0.02)	0.01

Participants who took antibiotics, pre/probiotics, or had diarrhea, irritable bowel syndrome, or inflammatory bowel disease were additionally excluded from the analysis. The mean (SD) for BMI (kg/m^2) and WHtR was 24.01 (3.15) and 0.52 (0.06), respectively. Because the SCFAs abundance were log_2 transformed, the linear model coefficients are interpreted as units of BMI and WHtR associated with a fold increase in SCFAs. Model was adjusted for age, sex, batch, province, urbanization, income, education, total energy, insoluble fiber, physical activity, smoking, and alcohol intake.

Table S6.11. Sensitivity analysis of the interaction between dietary precursors of short-chain fatty acids (SCFAs) and overweight in models of SCFAs [% change (95% CI)], corresponding to Figure S6.2[1]

n=462	Butyrate/isobutyrate	Valerate	Isovalerate	Total SCFAs
		Insoluble fiber (ref. = low)[2]		
Middle	-14.03 (-27.72, 2.25)	-12.85 (-41.67, 30.19)	-15.74 (-32.15, 4.63)	-14.27 (-28.07, 2.18)
High	0.26 (-16.23, 20)	-13.04 (-42.62, 31.79)	-0.24 (-20.29, 24.85)	0.75 (-15.99, 20.84)
Middle X High BMI	23.97 (-1.94, 56.71)	15.05 (-33.12, 97.89)	28.73 (-3.93, 72.5)	23.9 (-2.26, 57.05)
High X High BMI	1.45 (-19.72, 28.19)	38.81 (-19.22, 138.53)	4.52 (-21.96, 39.98)	7.43 (-15.21, 36.12)
Interaction p-value[3]	0.133	0.491	0.196	0.197
		Carbohydrate (ref. = low)[2]		
Middle	5.21 (-11.46, 25.01)	15.22 (-22.55, 71.4)	5.34 (-15.05, 30.61)	5.28 (-11.52, 25.27)
High	0.26 (-17.8, 22.28)	-11.79 (-44.16, 39.36)	-5.32 (-26.09, 21.28)	-4.18 (-21.56, 17.06)
Middle X High BMI	-1.03 (-21.74, 25.17)	-26.29 (-57.08, 26.59)	-1.48 (-26.49, 32.04)	-3.67 (-23.98, 22.06)
High X High BMI	9.74 (-13.39, 39.05)	17.45 (-31.91, 102.59)	15 (-14.39, 54.5)	18.63 (-6.55, 50.6)
Interaction p-value[3]	0.645	0.238	0.530	0.194
		Fiber foods (ref. = low)[2]		
Middle	-4.88 (-19.87, 12.91)	-6 (-36.74, 39.65)	-5.74 (-23.85, 16.68)	-7.72 (-22.37, 9.69)
High	-3.21 (-18.95, 15.58)	20.39 (-20.08, 81.36)	-0.83 (-20.48, 23.68)	-0.51 (-16.81, 18.98)

Middle X High BMI	-9.08 (-27.98, 14.79)	-6.31 (-45.3, 60.49)	-15.5 (-36.78, 12.94)	-9.6 (-28.53, 14.34)
High X High BMI	-5.21 (-25.03, 19.83)	-22.43 (-54.86, 33.31)	-6.27 (-30, 25.49)	-7.8 (-27.21, 16.79)
Interaction p-value[3]	0.724	0.635	0.517	0.670

[1] Participants who took antibiotics, pre/probiotics, or had diarrhea, irritable bowel syndrome, or inflammatory bowel disease were additionally excluded from the analysis. BMI$\geqslant$24 kg/m^2. The abundance of each SCFAs and the sum of the all three SCFAs were $\log_2$ transformed. The % change relative to the reference level was calculated using the following formula: (2^(β)-1)*100%, where β was the linear model coefficient. Model was adjusted for age, sex, batch, province, urbanization, income, education, total energy, insoluble fiber, physical activity, smoking, and alcohol intake. *, % change p-value<0.05, **, % change p-value<0.01.

[2] Dietary intakes were measured by 3-consecutive 24h dietary recalls and household food inventories. High-fiber foods was calculated as the sum of whole grains, legumes, starchy roots, vegetables, mushrooms/seaweeds, fruits, nuts/seeds. Insoluble fiber, carbohydrate, and high-fiber food score were categorized by tertiles to represent low, middle, and high intakes.

[3] The statistical significance of the interaction term between dietary factors and overweight was estimated using a Wald test that compared models with and without the interaction term.

Table S6.12. Sensitivity analysis of the interaction between dietary precursors of short-chain fatty acids (SCFAs) and abdominal obesity in models of SCFAs [% change (95% CI)], corresponding to Figure S6.3[1]

n=460	Butyrate/isobutyrate	Valerate	Isovalerate	Total SCFAs
		Insoluble fiber (ref. = low)[2]		
Middle	-18.6 (-33.56, -0.27)	-16.01 (-47.74, 34.99)	-19.59 (-37.68, 3.76)	-19.36 (-34.38, -0.9)*
High	0.7 (-18.25, 24.06)	-13.12 (-46.63, 41.44)	1.15 (-22.15, 31.42)	0.3 (-18.84, 23.95)
Middle X High WHtR	26.33 (-0.89, 61.04)	16.67 (-33.83, 105.71)	28.24 (-5.44, 73.92)	27.34 (-0.47, 62.92)
High X High WHtR	0.32 (-21.43, 28.09)	22.74 (-30.65, 117.24)	1.3 (-25.45, 37.67)	5.46 (-17.71, 35.15)
Interaction p-value[3]	0.099	0.762	0.201	0.132
		Carbohydrate (ref. = low)[2]		
Middle	2.77 (-16.34, 26.24)	-19.57 (-50.03, 29.46)	0.54 (-22.3, 30.11)	-5.69 (-23.46, 16.21)
High	13.75 (-9.6, 43.14)	-42.92 (-66.46, -2.87)*	12.45 (-15.68, 49.97)	-0.95 (-21.55, 25.07)
Middle X High WHtR	3.5 (-18.9, 32.08)	36.39 (-22.42, 139.77)	6.67 (-21.41, 44.79)	15.24 (-10.03, 47.6)
High X High WHtR	-10.79 (-30.57, 14.63)	110.58 (17.9, 276.14)*	-13.63 (-36.92, 18.25)	8.26 (-16.06, 39.63)
Interaction p-value[3]	0.473	0.042	0.396	0.530
		High-fiber foods (ref. = low)[2]		
Middle	-13.95 (-29.69, 5.31)	-1.38 (-38.39, 57.87)	-18.65 (-36.82, 4.75)	-15.49 (-31.12, 3.69)
High	-5.29 (-23.05, 16.58)	7.9 (-33.49, 75.06)	-2.38 (-24.73, 26.6)	-4.14 (-22.33, 18.3)

Middle X High WHtR	7.96 (-15.22, 37.47)	-10.78 (-49.19, 56.66)	9.64 (-18.98, 48.36)	5.64 (-17.29, 34.93)
High X High WHtR	-2.45 (-23.42, 24.28)	1.1 (-42.48, 77.7)	-4.53 (-29.49, 29.26)	-1.67 (-23.05, 25.65)
Interaction p-value[3]	0.701	0.893	0.669	0.840

[1] Participants who took antibiotics, pre/probiotics, or had diarrhea, irritable bowel syndrome, or inflammatory bowel disease were additionally excluded from the analysis. Waist-to-height ratio$\geqslant$0.5. The abundance of each SCFAs and the sum of the all three SCFAs were $\log_2$ transformed. The % change relative to the reference level was calculated using the following formula: (2^(β)-1)*100%, where β was the linear model coefficient. Model was adjusted for age, sex, batch, province, urbanization, income, education, total energy, insoluble fiber, physical activity, smoking, and alcohol intake. *, % change p-value<0.05, **, % change p-value<0.01.

[2] Dietary intakes were measured by 3-consecutive 24h dietary recalls and household food inventories. High-fiber foods was calculated as the sum of whole grains, legumes, starchy roots, vegetables, mushrooms/seaweeds, fruits, nuts/seeds. Insoluble fiber, carbohydrate, and high-fiber food score were categorized by tertiles to represent low, middle, and high intakes.

[3] The statistical significance of the interaction term between dietary factors and abdominal obesity was estimated using a Wald test that compared models with and without the interaction term.

Table S6.13. Sensitivity analysis of the association between the overall and the total relative abundance of 56 microbial short-chain fatty acid (SCFA) producers with body mass index (BMI) and waist-to-height-ratio (WHtR)

			Overall[1]		**Total[2]**	
	n	**Mean (SD)**	**R^2**	**P-value**	**β (95% CI)**	**P-value**
BMI	192	24.38 (3.12)	0.008	0.13	-0.95 (-2.82, 0.93)	0.32
WHtR	191	0.53 (0.06)	0.007	0.20	-0.03 (-0.06, 0.01)	0.14

Participants who took antibiotics, pre/probiotics, or had diarrhea, irritable bowel syndrome, or inflammatory bowel disease were additionally excluded from the analysis. The raw counts of each species and the total counts of the 56 species were normalized and $\log_{10}$ transformed [22]. Model was adjusted for age, sex, batch, province, urbanization, income, education, total energy, insoluble fiber, physical activity, smoking, and alcohol intake.

[1] R^2 and p-value were calculated using permutational multivariate analysis of variance (PERMANOVA) of all 56 species.

[2] Linear regression was performed on the total relative abundance of the 56 species.

Figure S6.1. Sample flow chart.

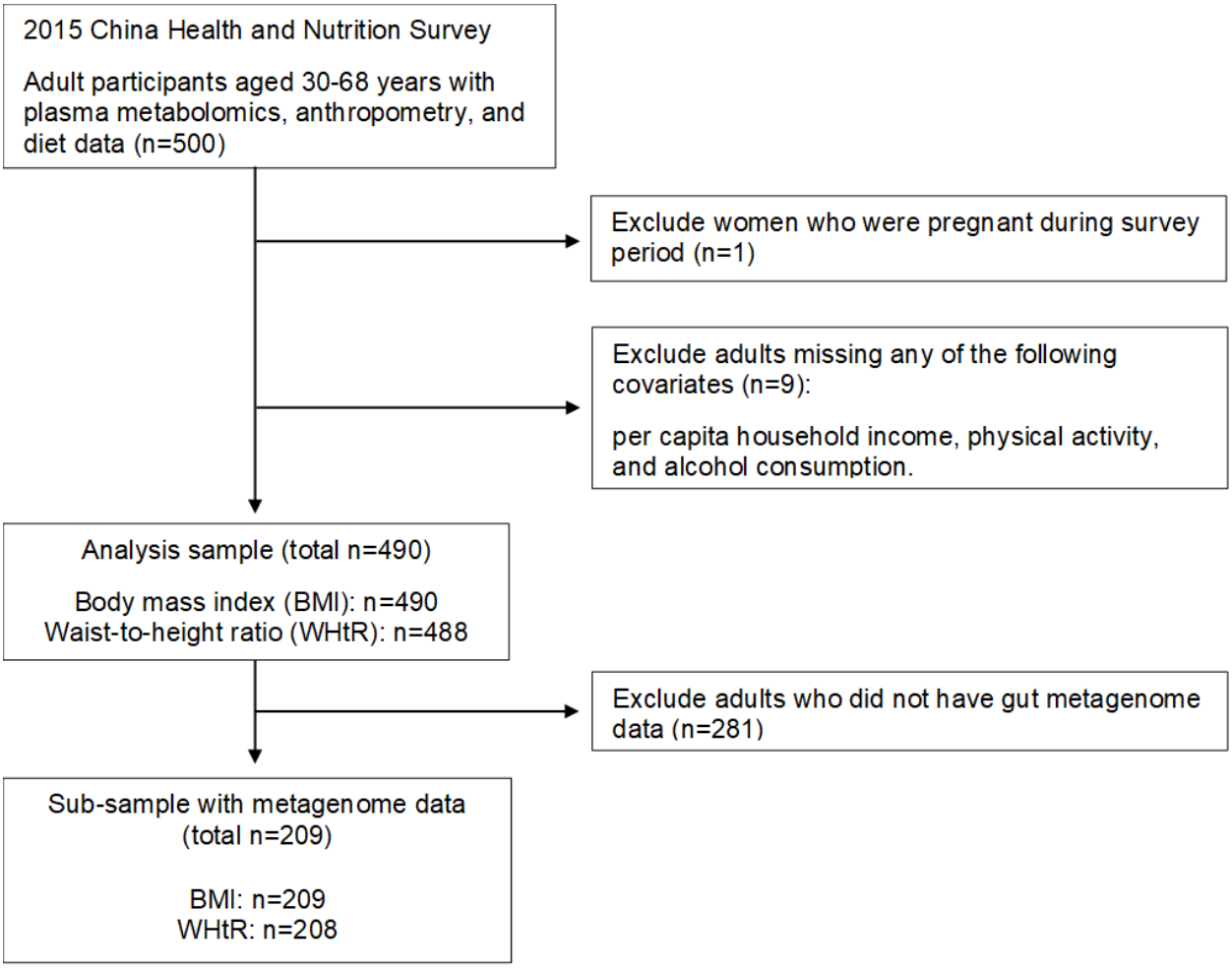


Figure S6.2. Sensitivity analysis of the associations between dietary precursors of short-chain fatty acids (SCFAs) and plasma SCFAs by overweight. Overweight: BMI⩾24 kg/m^2. Vertical axes represent model predicted (marginal means) SCFAs abundance. Dietary intakes were categorized by tertiles to represent low, middle, and high intakes. Linear model was adjusted for age, sex, batch run, province, urbanization, income, education, physical activity, total energy intake, insoluble fiber, alcohol, and ever smoking. P-value for the interaction between each dietary precursor of SCFAs and overweight was derived using a Wald test. P-value>0.05 for all comparisons of plasma SCFA abundance at a given level of a dietary precursor by overweight.

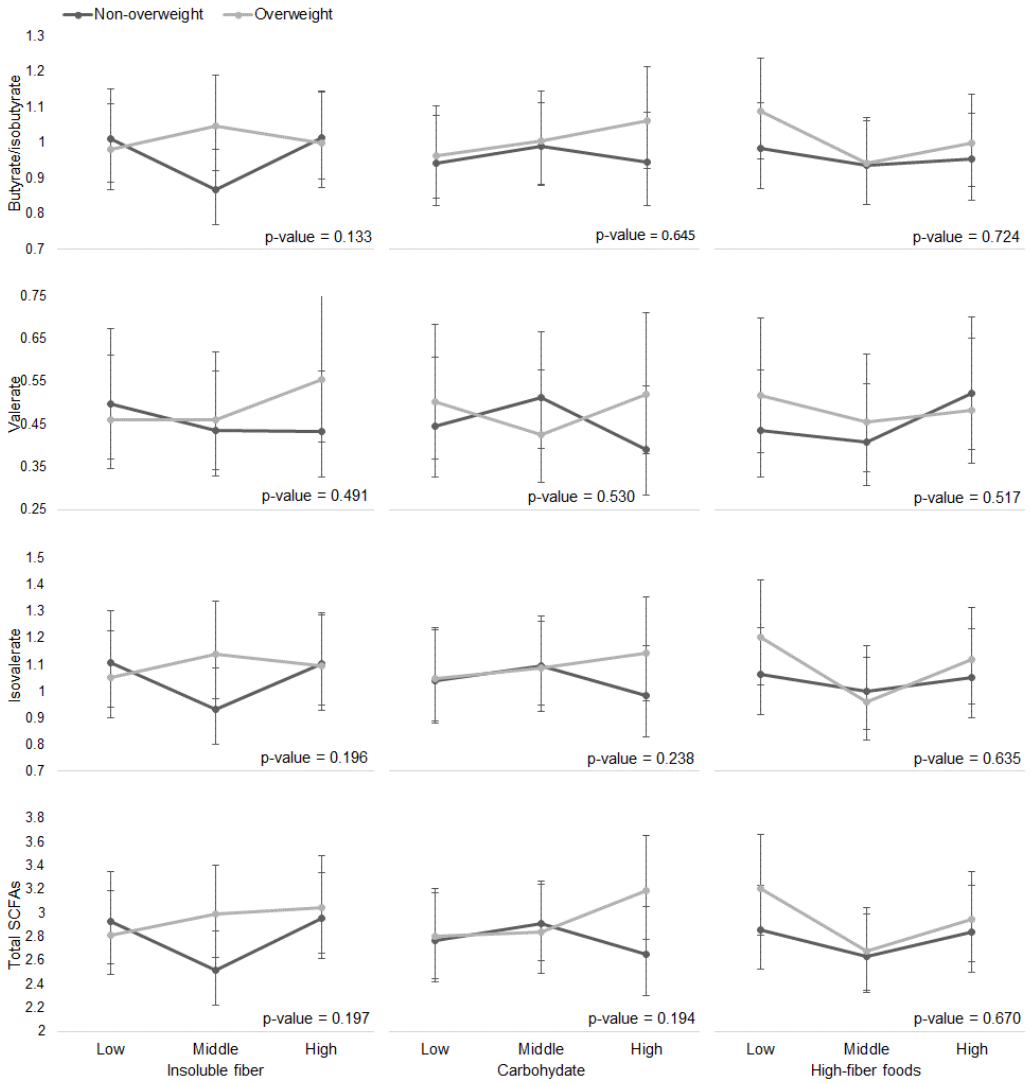
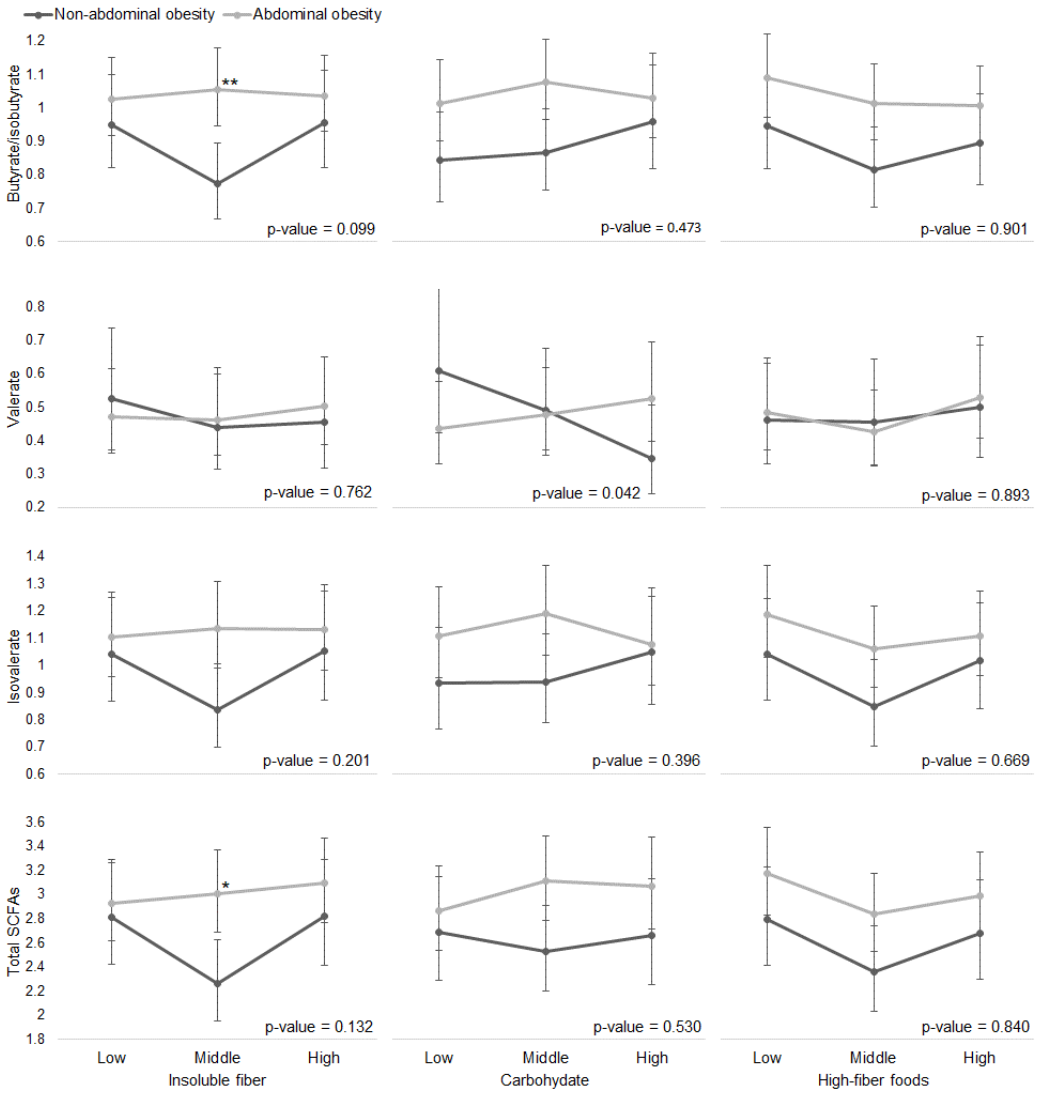


Figure S6.3. Sensitivity analysis of the associations between dietary precursors of short-chain fatty acids (SCFAs) with plasma SCFAs by abdominal obesity. Abdominal obesity: waist-to-height ratio⩾0.5. Vertical axes represent model predicted (marginal means) SCFAs abundance. Dietary intakes were categorized by tertiles. Linear model was adjusted for age, sex, batch run, province, urbanization, income, education, physical activity, total energy intake, insoluble fiber intake, alcohol, and ever smoking. P-value for the interaction between each dietary precursor of SCFAs and AOB was derived using a Wald test. *, p-value<0.5; **, p-value <0.01 for comparisons of plasma SCFAs abundance at a given level of intake by abdominal obesity



CHAPTER 7. SYNTHESIS

Overview of findings

The overall goal of this research was to understand the microbial and metabolic pathways underlying the dietary etiology of cardiovascular disease (CVD) risk factors. We also aimed to investigate these complex associations and provide insights into potential biomarkers and therapeutic targets to improve interventions and treatments. To achieve these purposes, we investigated the relationships between diet, gut microbiota, circulating metabolites, and CVD risk factors in free-living adults using well-characterized data from the China Health and Nutrition Survey (CHNS), which provided gut microbiota data across 12 provinces and three megacities and paired plasma metabolite data from four adjacent southern provinces. We also had access to high-quality diet data obtained from both three validated 24-h diet recalls and household food inventories, CVD risk factors assessed by trained clinicians, and detailed data of sociodemographic and health behavioral factors.

Our research addressed two major gaps in current literature of microbiota and metabolomics, the lack of strong population-based evidence and the lack of integrated analysis of microbiota and metabolites. We first showed gut microbiota and plasma metabolome were associated with sodium and potassium consumption, two key dietary risk factors for CVD, after accounting for geographic variation in diet and microbial communities. Second, we identified biologically plausible patterns of metabolites and assessed the associations between these patterns and the individual metabolites underlying these patterns with blood pressure in middle-aged adults during critical period for CVD development. We also examined the association between the overall gut microbial community and specific microbial groups with blood pressure. Last, we quantified the association between circulating short-chain fatty acids

(SCFAs), which have been shown to be involved in multiple pathways that modulate CVD risk factors, with body mass and abdominal adiposity measures. We additionally examined the roles of diet and microbiota in the SCFA-adiposity relationship. In this section, we provide a brief summary and synthesis of our findings.

Are gut microbiota and host metabolites involved in dietary sodium and potassium-associated CVD risk?

We examined associations between sodium and potassium consumption with gut microbiota and host metabolome. We analyzed of the overall microbial community and metabolome (distance-based redundancy analysis, dbRDA), as well as specific taxa and metabolites (multivariable-adjusted linear regression). To understand whether the associations between dietary sodium and potassium with host metabolome were potentially mediated by microbiota, we additionally examined the associations between specific taxa and metabolites that have shown associations with sodium and/or potassium consumption.

After accounting for geographic variation, we found that energy-adjusted sodium, potassium, and Na/K ratio consumption were associated with between-person gut microbial diversity and several specific genera. While in a few provinces, sodium tended to be positively associated with infectious pathogenic bacteria and genera shown to be linked to hypertension and obesity in the literature, potassium tended to be positively associated with SCFAs-producing commensals. These diet factors were also associated with the overall metabolome and several individua metabolites. Specifically, sodium consumption was negatively associated with the anti-inflammatory microbial-derived phenolics, while positively associated with SCFAs and pro-inflammatory fibrinogen cleavage peptide and eicosanoids, all of which are involved in CVD risk development. The positive associations between a few taxa with phenolics indicated that microbiota are indeed involved in the association between dietary sodium and microbiota-mediated metabolites. In addition, as energy-adjusted sodium, potassium, and Na/K ratio consumption were positively with several specific microbiota [both genus- and operational taxonomic unit (OTU)-level] and metabolites in some provinces and megacities, but negatively associated with them in other provinces and

megacities, our results suggested that geographic variation should be considered in future analyses. Investigating geographic variations in microbiota and metabolites may improve the understanding of regional- and inter-individual differences in CVD risk attributable to high sodium and low potassium consumption [95]. Altogether, our results indicate that gut microbiota and circulating metabolites may play important roles in the dietary etiology of CVD and suggest potential routes of diet intake in relation to health outcomes. Thus, gut microbiota and related metabolites could be potential targets for sodium and potassium intervention to curb the CVD epidemic.

What are the underlying microbial and metabolic pathways of blood pressure regulation?

By examining the association between gut microbiota and plasma metabolites with blood pressure, we aimed to test whether microbiota and metabolites that have been shown to regulate blood pressure in animal models, like SCFAs [80] and glycerrhetinic acid-like factors (GALFs) [28,33], were associated with blood pressure in a free-living general adult population. We analyzed the overall microbial community (permutational multivariate ANOVA) and specific genera and metabolites (multivariable-adjusted linear regression), as well as metabolite patterns derived from principal component analysis.

We found that blood pressure varied across between-person gut microbial diversity assessed by principal coordinate analysis (PCoA), with the fourth dimension of dissimilarity matrix positively associated with systolic blood pressure. A lipid pattern was positively associated with blood pressure, independent of estimated glomerular filtration rate (eGFR) and BMI, a potential mediator. Individual metabolites contributing to this pattern (e.g., acyl-carnitines and long-chain polyunsaturated and saturated fatty acids) and a few sphingomyelins and phosphatidylinositol were also positively associated with blood pressure. In random forest models, host sociodemographic and behavioral factor, microbiota, and metabolite data had comparable accuracy in predicting blood pressure, indicating that these data are equally valuable in studying blood pressure. Our findings also suggest that instead of a few specific

microbes, the overall microbiome may be involved in blood pressure regulation. Thus, interventions may need to modulate the whole microbial community to modulate high blood pressure. Similarly, a collection of circulating lipids, especially sphingomyelins, acyl-carnitine, and long-chain fatty acids, may need to be strictly monitored. Although we did not fully replicate the findings from studies using animal models (i.e., no association found for SCFAs and GALFs), our results shed light on effective management of high blood pressure targeting microbiota and specific groups of metabolites.

Circulating SCFAs and adiposity: How do diet and gut microbiota play a part?

We examined the association between individual plasma SCFAs and the total abundance of SCFAs with two adiposity measures, BMI and waist-to-height ratio (WHtR), using multivariable adjusted linear regression. To test whether subjects with higher adiposity had higher capacity to harvest energy through gut microbial production of SCFAs from carbohydrate fermentation, we first assessed whether the associations between dietary precursors of SCFAs and plasma SCFAs varied across levels of BMI and WHtR, and then we tested the associations between bacterial species shown to produce SCFAs with BMI and WHtR.

We found positive associations between SCFAs with BMI and WHtR, including butyrate/isobutyrate, isovalerate, and the total SCFAs. The associations between insoluble fiber with butyrate/isovalerate and total SCFAs and between total carbohydrate and valerate were slightly differed by WHtR levels. We also observed a differential association between insoluble fiber and butyrate/isovalerate across levels of BMI. For example, whereas valerate decreased as total carbohydrate intake increased in adults with normal WHtR, valerate tended to increase as total carbohydrate intake increased in adults with high WHtR. There was no association between the total relative abundance of SCFA-producing bacteria with BMI and WHtR. Our results indicate that the higher SCFAs associated with higher adiposity may reflect an interplay of diet and gut microbiome. Our results also inform future

design of dietary interventions for obesity the importance of considering potential inter-individual differences in the abilities of gut microbiota to digest dietary carbohydrates.

Strength and limitations

Limitations

One of the biggest challenges of the current research is the cross-sectional design. We were unable to establish the temporalities or causal-relationships between diet, microbiota, metabolites, and CVD risk factors. It is possible that the observed associations were due to reverse causation. That is, instead of leading to increased CVD risk, certain microbiota or metabolites may reflect the increased CVD risk. There are a few ways to infer causality in cross-sectional studies, for example, Mendelian randomization [203], which uses the host genetics as an instrumental variable for a single or a group of microbiota or metabolite. However, due to weak polygenetic score instrument for metabolites (i.e., low explained variance $R^2 \leq 2\%$, small sample size), we were unable to assess the causality using this approach. In addition, we were unable to test the stability of microbiota and metabolite data given the lack of repeated measures. There may be large within-individual variations in omics data, especially for circulating metabolites. As such, associations based on a single measure of metabolites could be biased when the intra class correlation<0.5, which indicates that the within-individual variance is large relative to the between-individual variance [139]. To reduce the variability in metabolites, we have restricted the metabolomics sample to middle aged adults from two adjacent southern provinces with similar customs and collected fasting blood samples at the same day per examination center following the same standardized protocol. In addition, we have previously reported that the CHNS gut microbiota samples were stable over two weeks [119].

Another challenge of our research is the lack of an independent replication sample. Many of our analyses were hypothesis-generating steps to delineate various biological pathways associated with diet and CVD risk factors. These analyses need to be repeated in other cohorts in order to confirm our findings

and inform the next steps in establishing causal roles of these pathways. Our exclusion criteria including the antibiotic use were based on *a priori* knowledge that they may affect the gut microbiome [53]. However, it is possible that these exclusions may create systematic differences in characteristics including socioeconomic status and health consciousness, making the analysis sample less representative of the general population and thus may leading to deviated results from the future replication cohort. Moreover, given that our unique Chinese cohort may have different dietary habits and dietary sources than other race/ethnic groups, for example, the extremely high sodium and low potassium intake, we acknowledge that our findings may have limited generalizability to other populations, like those with lower sodium and higher potassium intake than our cohort. Despite that the majority of our sample consumed excessed sodium and deficient potassium, we were still able to examine large variation in sodium and potassium consumption.

We cannot exclude the possibility of measurement errors, especially for diet assessment tools. Measurement errors in dietary intakes of sodium are common across studies [204] and the use of 24-h recall has shown less accurate estimation of sodium intake than 24-h urine in China [138]. Our use of both three-consecutive 24-h recalls and three-day household inventories, validated by three-consecutive 24-h urine, reduced these potential measurement errors. Nevertheless, in analysis of fiber intake, our diet assessment tools did distinguish fibers types, preventing us from conducting more detailed analysis of fermentable versus non-fermentable fibers. Measurement errors in covariates may exists as well and could lead to residual confounding. For example, we were unable to adjust for more detailed classification of smoking like former/current smoking due to small numbers of former smokers, though our definition of never/ever smokers was based on repeated measures in previous survey rounds to reduce miss classification.

An inherent limitation of 16S rRNA data is that it only provides microbial community structures, and not specific pathways based on microbial functional genes. We were unable to infer microbial and metabolic pathways directly from our 16S rRNA results. Though we have gut metagenome data in the

sub-sample, which aided us in identifying specific bacteria species that produce SCFAs, we may have limited statistical power owing to the small sample size (n=213) and limited comparability to current literatures that primarily based on genus-level 16S data. While we had similar or larger sample size than previous studies, we may still lack statistical power to test numerous potential interactions in our analyses (i.e., variations in geographic locations and adiposity levels). Because we examined thousands of specific taxa and metabolites, it is almost impossible to tailor our model for each taxa and metabolites. Our analyses were primarily based on linearity assumption and thus were preliminary steps to identify differential associations across provinces/megacities. It is a promising area for further regional-specific analysis to fully understand these variations.

Last, while the comprehensive snapshot of systemic metabolic processes captured by non-targeted metabolomics allowed us to test a wide range of metabolites, we did not examine metabolite concentrations or measure all metabolites of our interest. For instance, acetate and propionate with molecular sizes below detection level could not be identified in our metabolomics analysis. We also lacked fecal and urine metabolomics paired to circulating metabolomics to assess the absorption and excretion of microbial-mediated metabolites, which are important to understand the full mechanisms underlying the associations across diet, microbiota, metabolite, and CVD risk factors.

Strengths

Despite these limitations, our research has several strengths, including the well-characterized, and population-based cohort, with participants across 12 provinces and megacities and from a range of urban and rural communities, allowing us to detect large variations in diet and health factors. Given our diverse data, our findings are potentially more generalizable to the Chinese adult population than studies using data from single or a few communities. Our previous work has shown the gut microbiota varied substantially by provinces and megacities [118] and by urban versus rural areas [119], and thus sampling from diverse geographic locations is important for gut microbiome research. Additionally, our unique

cohort with low treatment rate for blood pressure enabled us to assess large variations in untreated blood pressure and ensured minimal medication effects to associations across microbiota, metabolites and blood pressure.

The CHNS provided high quality individual-, household-, and community-level data, spanning from self-reported sociodemographic information and health activities, clinically measured biomarkers, blood pressure and anthropometry, to community services, infrastructure and environment collected from officials, informants and official records. These rich data allowed use to control for an array of potential confounders, including urbanization, education, income, physical activity, smoking, alcohol intake, and eGFR. In particular, our diet measurement instruments, three-consecutive 24-h recall and household food inventories, had been validated by double labeled water for total energy intake [100] and three-consecutive 24-h urine excretion for sodium and potassium [95]. Our diet data captured cooking methods and all foods consumed within the household, as well as food consumed away from home. Our protocols for collections and processing of fecal samples and fasting blood samples were also standardized with strict quality control.

We had access to paired gut microbiota and plasma metabolomics data measured concurrently, allowing us to infer host metabolic pathways beyond microbial pathways and conduct integrated omics analysis, which further improved our understanding of how diet and microbiota may jointly affect metabolites. This feature also enabled us to compare microbiota and metabolite data to provide insight to which data had stronger association with or better prediction for host diet and physiology, compared to conventionally measured host sociodemographic and behavioral data. To our knowledge, few population-based studies have reported diet and health factors associated with microbiome and metabolome data.

Furthermore, we used multiple methodological approaches to more comprehensively analyze the unique characteristics of each types of data, which might otherwise be overlooked using a single approach. For example, we used partial dbRDA to simultaneously test the statistical significance of the main effect (i.e., sodium and potassium consumption) and the interaction effect (i.e., interaction of

sodium or potassium with provinces and megacities) on the overall microbial community and metabolome, which could not be done using the widely-used permutational multivariate ANOVA. This joint test is common in studies testing gene-environment interactions and has been shown to offer more statistical power than other methods when interaction exists and similar power to analysis of main effect only in the absence of an interaction [124]. We also used principal component analysis (PCA) to extract biological possible metabolite patterns given the intricate correlations across metabolites, which cannot be accounted for if metabolites were only examined individually. Finally, the use of random forest regression, a machine learning approach, allowed us to compare across the whole set of data for microbiota, metabolites, and other host factors in association with diet and CVD risk factors.

Public health significance

Our research has critical implications for improving public health. The prevalence of leading CVD risk factors like high blood pressure and overweight/obesity continue to rise around the world [6,15,16]. Research on gut microbiota and host metabolites could provide valuable insights into the multifactorial systems that regulate blood pressure and body weight, therefore shedding light on effective personalized interventions and treatment that ultimately help advancing population health. Findings from the current research can also inform future studies about specific hypotheses on relationships across diet, microbiota, metabolites and CVD risk factors.

Our results add evidence to the potential deleterious health effects of high sodium and low potassium consumption

We found that dietary sodium and potassium were positively and negatively associated with several pathogenic bacteria like *Staphylcoccus, Pseudomonas,* and *Moraxellaceae*, respectively, across a few provinces and megacities in our sample. Sodium was also negatively associated with the anti-inflammatory gut-derived phenolics, while positively associated with pro-inflammatory fibrinogen

cleavage peptide and eicosanoid in Guizhou province. Our findings are in line with mouse models indicating that the increased inflammation plays a part in high sodium-induced high blood pressure [42]. Sodium-induced colon inflammation also exacerbated colitis in mice [136]. Therefore, our research supports the urgent needs to increase health consciousness of dietary sodium and potassium intakes. Our findings also suggest a possibility to reverse the harmful effects of high sodium and low potassium intakes on health by modulating the gut microbiota. For example, *Lactobacillus* supplementation reduced blood pressure in mice with high blood pressure induced by high sodium intake [42]. The substantial variations in our results across geographic locations may reflect different basal microbiome and different dietary sources of sodium and potassium, indicating that dietary interventions tailored for each region or community may be more appropriate than a non-tailored intervention, though more studies are needed to confirm this.

Our findings unravel potential biological pathways underlying high blood pressure

In our research, a lipids pattern (e.g., long-chain fatty acids), several individual lipid metabolites, and a few metabolic pathways (e.g., sphingolipids, long-chain acyl-carnitines) were positively associated with blood pressure. Others have shown that the metabolomics-related mechanisms may impact blood pressure. For instance, excess acyl-carnitines can stimulate proinflammatory pathways through Nuclear factor kappa B (NF-κB) [161], and long-chain fatty acids like linoleate can be metabolized to arachidonic acid, a precursor of proinflammatory eicosanoids [162]. Moreover, sphingolipids are candidate biomarkers for CVD and can predict coronary artery disease better than LDL-cholesterol and triglycerides [165-167]. Thus, our results suggest that these blood lipids may need to be measured in combination with other traditional blood markers to thoroughly evaluate individual and population risk for CVD. These lipids may also be important targets for CVD prevention and treatment.

Our results also provide insights into an interplay of diet and gut microbiota in high adiposity

We found positive associations between plasma SCFAs and adiposity measures. Studies have suggested that SCFAs are a large source of dietary energy extracted by the gut microbiota, adding as much as 10% extra energy to adults [86]. Others have also shown that the gut microbiota of people with obesity had higher capacity to harvest energy, as measured by SCFA production after fermenting the same amount of carbohydrate, than those with normal weight [73,179]. Goffredo et al. [73] showed that the microbiota of adolescents with higher body fat was characterized by higher *Firmicutes/Bacteroidete*s ratio and specific genera including *Streptococcus*, *Actinomyces* and *Blautia.* These results suggest that the positive associations between SCFAs and adiposity are at least partially due to the higher energy harvesting in obese versus lean gut microbiota. In other words, even when consuming the same amount of carbohydrate, people with an obese microbial phenotype may potentially obtain more energy than people with a lean-associated microbiota. As such, our findings indicate that the gut microbiota may be a modifiable target employed with dietary interventions for obesity.

In addition, our results suggest that high-fiber foods like fruits, nuts/seeds, and whole grains may be beneficial for reducing abdominal adiposity by potentially helping decrease energy harvesting as indicated by circulating SCFAs. We have shown that these high-fiber foods were negatively associated with plasma SCFAs in adults with high WHtR. Similar results were found in a double-blinded, randomized-controlled human trial, in which the prebiotic fiber groups had lower fecal SCFAs than the placebo group after treatment period [180]. Thus, our findings shed light on the potentiality to monitoring microbiota and microbiota-mediated metabolites using diet to decrease one's susceptibility to accumulate adiposity. Overall, our research highlights the importance of considering diet-microbiota interaction as candidates for CVD risk factor management.

Future directions

There are many research directions that could be extended from our analyses to further advance our understanding the roles of gut microbiota and circulating metabolites in the diet-CVD risk relationships.

A pivotal objective of future research is to establish the temporal relationships across diet, microbiota, metabolites, and CVD risk factors. Population-based cohorts with repeated measures of microbiota and metabolomics data are needed to determine the longitudinal associations between diet with microbiota and metabolites and to identify novel microbiota and metabolite markers predicting elevated CVD risk. Our future research using the 2015 and the upcoming round of CHNS will allow us to investigate these associations. Ideally, using more reliable dietary measures like doubly-labeled water for total energy and urine excretion for sodium and potassium than self-reported data may further improve the research. Host genotype could also be employed to infer temporality, as it can act as an instrument variable for the exposure, a strategy known as Mendelian randomization [6]. In theory, host genotype is unlikely be influenced by the outcome of interest and environment and lifestyle, which are common confounding factors of CVD epidemiology, thereby limiting reverse causality and confounding in observational studies.

Randomized-control trials (RCTs) are essential to clarify how a particular diet affects CVD risk through alterations in microbiota and metabolites. There are a few cross-over trials examined the association between sodium restriction and changes in circulating metabolomics, such as the Dietary Approaches to Stop Hypertension (DASH)-Sodium trial [67]. Yet, to date, few large population-representative RCTs have examined the effects of sodium or other key dietary risk factors for CVD like potassium and fiber on both microbiota and metabolomics. Because we observed negative associations between fiber-rich foods and SCFAs in adults with high abdominal adiposity, it is of our interest to conduct RCTs to examine whether a high fiber diet reduced obesity by shifting the gut microbiota from higher to lower energy-harvesting capacity and whether this effect varies by different types of fiber. It is

also of our interest to test whether gut microbiota and metabolites explain or reflect individual response to diet, like salt sensitivity.

Future research could add analyses for gut microbial functional genes using the metagenomics data to examine the specific microbial pathways influenced by diet and/or affecting CVD risk. Specifically, these analyses could help understand why the association between dietary sodium and potassium with gut microbiota varied substantially across geographic locations in our sample. These analyses could also help us identify groups of microbes with similar functions to test specific hypotheses. For example, metagenome analysis allows identification of a full list of bacteria exhibiting nitrogen monoxide- and hydrogen sulfide- producing pathways, which are two gases that have been shown to reduce blood pressure [205,206]. We can then examine what diet patterns or specific foods and nutrients affect these bacteria to inform future diet intervention.

In addition, based on our results and prior research, several specific hypotheses could be tested, such as the microbial-mediated metabolites, GALFs and SCFAs that have established roles in high blood pressure [28,33,80]. These specific metabolites may require targeted measures as they may be poorly captured by nontargeted metabolomics, especially acetate and propionate with small molecular sizes. Multiple types of specimens, including fecal, blood, and urine, may also be needed to better capture transient metabolites in circulation, like butyrate [159], and to fully understand the association between the production, absorption, and excretion of these metabolites with CVD risk factors. Finally, while we only examined three major CVD risk factors (i.e., blood pressure, body mass and abdominal circumference), our analyses could be extended to many other risk factors. For instance, the metabolite patterns derived from PCA in our research could be further assessed to investigate their associations with type 2 diabetes and high c-reaction protein.

In conclusion, gut microbiota and host metabolome reveal the complex mechanisms linking diet with CVD risk factors. Our research has identified a few potential routes of action, including excess blood lipids, and provided important insights into future studies. As such, microbiota and metabolites could be

promising biomarkers for CVD dietary interventions. However, research in this area is still in its infancy. Admittedly, it may take years to develop strong interventions and therapy targeting these candidate markers based on our current and future research. In general, more studies are needed to fully elucidate the associations and interplays across diet, microbiota, metabolites, and CVD risk factors.

REFERENCES

1. Mensah, G.A.; Roth, G.A.; Fuster, V. The Global Burden of Cardiovascular Diseases and Risk Factors: 2020 and Beyond. Journal of the American College of Cardiology: 2019.

2. Benjamin, E.J.; Muntner, P.; Bittencourt, M.S. Heart disease and stroke statistics-2019 update: a report from the American Heart Association. *Circulation* **2019**, *139*, e56-e528.

3. Twig, G.; Yaniv, G.; Levine, H.; Leiba, A.; Goldberger, N.; Derazne, E.; Ben-Ami Shor, D.; Tzur, D.; Afek, A.; Shamiss, A. Body-mass index in 2.3 million adolescents and cardiovascular death in adulthood. *New England Journal of Medicine* **2016**, *374*, 2430-2440.

4. Khan, S.S.; Ning, H.; Wilkins, J.T.; Allen, N.; Carnethon, M.; Berry, J.D.; Sweis, R.N.; Lloyd-Jones, D.M. Association of body mass index with lifetime risk of cardiovascular disease and compression of morbidity. *JAMA cardiology* **2018**, *3*, 280-287.

5. Finkelstein, E.A.; Trogdon, J.G.; Cohen, J.W.; Dietz, W. Annual Medical Spending Attributable To Obesity: Payer-And Service-Specific Estimates: Amid calls for health reform, real cost savings are more likely to be achieved through reducing obesity and related risk factors. *Health affairs* **2009**, *28*, w822-w831.

6. Ng, M.; Fleming, T.; Robinson, M.; Thomson, B.; Graetz, N.; Margono, C.; Mullany, E.C.; Biryukov, S.; Abbafati, C.; Abera, S.F. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The lancet* **2014**, *384*, 766-781.

7. Ogden, C.L.; Carroll, M.D.; Fryar, C.D.; Flegal, K.M. Prevalence of obesity among adults and youth: United States, 2011–2014. **2015**.

8. Huffman, M.D.; Capewell, S.; Ning, H.; Shay, C.M.; Ford, E.S.; Lloyd-Jones, D.M. Cardiovascular health behavior and health factor changes (1988–2008) and projections to 2020: results from the National Health and Nutrition Examination Surveys. *Circulation* **2012**, *125*, 2595-2602.

9. Ashwell, M.; Gunn, P.; Gibson, S. Waist-to-height ratio is a better screening tool than waist circumference and BMI for adult cardiometabolic risk factors: systematic review and meta-analysis. *Obesity reviews* **2012**, *13*, 275-286.

10. Albrecht, S.S.; Gordon-Larsen, P.; Stern, D.; Popkin, B.M. Is waist circumference per body mass index rising differentially across the United States, England, China and Mexico? *European journal of clinical nutrition* **2015**, *69*, 1306.

11. Patel, S.A.; Winkel, M.; Ali, M.K.; Narayan, K.V.; Mehta, N.K. Cardiovascular mortality associated with 5 leading risk factors: national and state preventable fractions estimated from survey data. *Annals of internal medicine* **2015**, *163*, 245-253.

12. Bromfield, S.; Muntner, P. High blood pressure: the leading global burden of disease risk factor and the need for worldwide prevention programs. *Current hypertension reports* **2013**, *15*, 134-136.

13. Wang, Z.; Chen, Z.; Zhang, L.; Wang, X.; Hao, G.; Zhang, Z.; Shao, L.; Tian, Y.; Dong, Y.; Zheng, C. Status of hypertension in China: results from the China Hypertension Survey, 2012–2015. *Circulation* **2018**, *137*, 2344-2356.

14. Li, D.; Lv, J.; Liu, F.; Liu, P.; Yang, X.; Feng, Y.; Chen, G.; Hao, M. Hypertension burden and control in mainland China: analysis of nationwide data 2003–2012. *International journal of cardiology* **2015**, *184*, 637-644.

15. Forouzanfar, M.H.; Liu, P.; Roth, G.A.; Ng, M.; Biryukov, S.; Marczak, L.; Alexander, L.; Estep, K.; Abate, K.H.; Akinyemiju, T.F. Global burden of hypertension and systolic blood pressure of at least 110 to 115 mm Hg, 1990-2015. *Jama* **2017**, *317*, 165-182.

16. Mills, K.T.; Bundy, J.D.; Kelly, T.N.; Reed, J.E.; Kearney, P.M.; Reynolds, K.; Chen, J.; He, J. Global burden of hypertension: analysis of population-based studies from 89 countries. *Journal of Hypertension* **2015**, *33*, e2.

17. Mozaffarian, D.; Fahimi, S.; Singh, G.M.; Micha, R.; Khatibzadeh, S.; Engell, R.E.; Lim, S.; Danaei, G.; Ezzati, M.; Powles, J. Global sodium consumption and death from cardiovascular causes. *New England Journal of Medicine* **2014**, *371*, 624-634.

18. Härtl, G. WHO issues new guidance on dietary

salt and potassium. Availabe online: https://www.who.int/mediacentre/news/notes/2013/salt_potassium_20130131/en/ (accessed on November).

19. Jackson, S.L.; Cogswell, M.E.; Zhao, L.; Terry, A.L.; Wang, C.-Y.; Wright, J.; Coleman King, S.M.; Bowman, B.; Chen, T.-C.; Merritt, R. Association between urinary sodium and potassium excretion and blood pressure among adults in the United States: National Health and Nutrition Examination Survey, 2014. *Circulation* **2018**, *137*, 237-246.

20. Yoon, Y.; Oh, S. Sodium density and obesity; the Korea National Health and Nutrition Examination Survey 2007–2010. *European journal of clinical nutrition* **2013**, *67*, 141-146.

21. Schwingshackl, L.; Schwedhelm, C.; Hoffmann, G.; Knüppel, S.; Iqbal, K.; Andriolo, V.; Bechthold, A.; Schlesinger, S.; Boeing, H. Food groups and risk of hypertension: a systematic review and dose-response meta-analysis of prospective studies. *Advances in Nutrition* **2017**, *8*, 793-803.

22. Cho, S.S.; Qi, L.; Fahey Jr, G.C.; Klurfeld, D.M. Consumption of cereal fiber, mixtures of whole grains and bran, and whole grains and risk reduction in type 2 diabetes, obesity, and cardiovascular disease. *The American journal of clinical nutrition* **2013**, *98*, 594-619.

23. McGill, C.R.; Victor III, L.F.; Devareddy, L. Ten-year trends in fiber and whole grain intakes and food sources for the United States population: National Health and Nutrition Examination Survey 2001–2010. *Nutrients* **2015**, *7*, 1119-1130.

24. He, J.; Gu, D.; Chen, J.; Jaquish, C.E.; Rao, D.C.; Hixson, J.E.; Chen, J.-c.; Duan, X.; Huang, J.-f.; Chen, C.-S. Gender difference in blood pressure responses to dietary sodium intervention in the GenSalt study. *Journal of hypertension* **2009**, *27*, 48.

25. Luft, F.C.; Weinberger, M.H. Heterogeneous responses to changes in dietary salt intake: the salt-sensitivity paradigm. *The American journal of clinical nutrition* **1997**, *65*, 612S-617S.

26. Morimoto, A.; Uzu, T.; Fujii, T.; Nishimura, M.; Kuroda, S.; Nakamura, S.; Inenaga, T.; Kimura, G. Sodium sensitivity and cardiovascular events in patients with essential hypertension. *The Lancet* **1997**, *350*, 1734-1737.

27. Weinberger, M.H.; Fineberg, N.S.; Fineberg, S.E.; Weinberger, M. Salt sensitivity, pulse pressure, and death in normal and hypertensive humans. *Hypertension* **2001**, *37*, 429-432.

28. Morris, D.J.; Ridlon, J.M. Glucocorticoids and gut bacteria:"The GALF Hypothesis" in the metagenomic era. *Steroids* **2017**, *125*, 1-13.

29. Dave, L.A.; Hayes, M.; Montoya, C.A.; Rutherfurd, S.M.; Moughan, P.J. Human gut endogenous proteins as a potential source of angiotensin-I-converting enzyme (ACE-I)-, renin inhibitory and antioxidant peptides. *Peptides* **2016**, *76*, 30-44.

30. Sircana, A.; De Michieli, F.; Parente, R.; Framarin, L.; Leone, N.; Berrutti, M.; Paschetta, E.; Bongiovanni, D.; Musso, G. Gut Microbiota, Hypertension and Chronic kidney Disease: recent advances. *Pharmacological research* **2018**.

31. Galla, S.; Chakraborty, S.; Mell, B.; Vijay-Kumar, M.; Joe, B. Microbiotal-host interactions and hypertension. *Physiology* **2017**, *32*, 224-233.

32. Feighner, S.D.; Hylemon, P.B. Characterization of a corticosteroid 21-dehydroxylase from the intestinal anaerobic bacterium, Eubacterium lentum. *Journal of lipid research* **1980**, *21*, 585-593.

33. Morris, D.; Latif, S.; Hardy, M.; Brem, A. Endogenous inhibitors (GALFs) of 11β-hydroxysteroid dehydrogenase isoforms 1 and 2: Derivatives of adrenally produced corticosterone and cortisol. *The Journal of steroid biochemistry and molecular biology* **2007**, *104*, 161-168.

34. Edwards, C.; Burt, D.; McIntyre, M.; De Kloet, E.; Stewart, P.; Brett, L.; Sutanto, W.; Monder, C. Localisation of 11β-hydroxysteroid dehydrogenase—tissue specific protector of the mineralocorticoid receptor. *The Lancet* **1988**, *332*, 986-989.

35. 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. Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **2013**, *341*, 1241214.

36. Adnan, S.; Nelson, J.W.; Ajami, N.J.; Venna, V.R.; Petrosino, J.F.; Bryan Jr, R.M.; Durgan, D.J. Alterations in the gut microbiota can elicit hypertension in rats. *Physiological genomics* **2016**, *49*, 96-104.

37. Gómez-Guzmán, M.; Toral, M.; Romero, M.; Jiménez, R.; Galindo, P.; Sánchez, M.; Zarzuelo, M.J.; Olivares, M.; Gálvez, J.; Duarte, J. Antihypertensive effects of probiotics Lactobacillus strains in spontaneously hypertensive rats. *Molecular nutrition & food research* **2015**, *59*, 2326-2336.

38. 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. A core gut microbiome in obese and lean twins. *nature* **2009**, *457*, 480-484.

39. Li, J.; Zhao, F.; Wang, Y.; Chen, J.; Tao, J.; Tian, G.; Wu, S.; Liu, W.; Cui, Q.; Geng, B. Gut microbiota dysbiosis contributes to the development of hypertension. *Microbiome* **2017**, *5*, 14.

40. Yang, T.; Santisteban, M.M.; Rodriguez, V.; Li, E.; Ahmari, N.; Carvajal, J.M.; Zadeh, M.; Gong, M.; Qi, Y.; Zubcevic, J. Gut dysbiosis is linked to hypertension. *Hypertension* **2015**, *65*, 1331-1340.

41. Yan, Q.; Gu, Y.; Li, X.; Yang, W.; Jia, L.; Chen, C.; Han, X.; Huang, Y.; Zhao, L.; Li, P. Alterations of the gut microbiome in hypertension. *Frontiers in cellular and infection microbiology* **2017**, *7*, 381.

42. Wilck, N.; Matus, M.G.; Kearney, S.M.; Olesen, S.W.; Forslund, K.; Bartolomaeus, H.; Haase, S.; Mähler, A.; Balogh, A.; Markó, L. Salt-responsive gut commensal modulates T H 17 axis and disease. *Nature* **2017**, *551*, 585.

43. Sun, S.; Lulla, A.; Sioda, M.; Winglee, K.; Wu, M.C.; Jacobs Jr, D.R.; Shikany, J.M.; Lloyd-Jones, D.M.; Launer, L.J.; Fodor, A.A. Gut Microbiota Composition and Blood Pressure: The CARDIA Study. *Hypertension* **2019**, *73*, 998-1006.

44. Tanaka, S.; Yoshida, M.; Murakami, Y.; Ogiwara, T.; Shoji, M.; Kobayashi, S.; Watanabe, S.; Machino, M.; Fujisawa, S. The relationship of Prevotella intermedia, Prevotella nigrescens and Prevotella melaninogenica in the supragingival plaque of children, caries and oral malodor. *Journal of Clinical Pediatric Dentistry* **2008**, *32*, 195-200.

45. Wang, X.; Yue, T.-L.; Barone, F.C.; White, R.F.; Clark, R.K.; Willette, R.N.; Sulpizio, A.C.; Aiyar, N.V.; Ruffolo, R.R.; Feuerstein, G.Z. Discovery of adrenomedullin in rat ischemic cortex and evidence for its role in exacerbating focal brain ischemic damage. *Proceedings of the National Academy of Sciences* **1995**, *92*, 11480-11484.

46. Podschun, R.; Ullmann, U. Klebsiella spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. *Clinical microbiology reviews* **1998**, *11*, 589-603.

47. Miquel, S.; Martin, R.; Rossi, O.; Bermudez-Humaran, L.; Chatel, J.; Sokol, H.; Thomas, M.; Wells, J.; Langella, P. Faecalibacterium prausnitzii and human intestinal health. *Current opinion in microbiology* **2013**, *16*, 255-261.

48. Rechner, A.R.; Smith, M.A.; Kuhnle, G.; Gibson, G.R.; Debnam, E.S.; Srai, S.K.S.; Moore, K.P.; Rice-Evans, C.A. Colonic metabolism of dietary polyphenols: influence of structure on microbial fermentation products. *Free Radical Biology and Medicine* **2004**, *36*, 212-225.

49. Zheng, Y.; Yu, B.; Alexander, D.; Mosley, T.H.; Heiss, G.; Nettleton, J.A.; Boerwinkle, E. Metabolomics and incident hypertension among blacks: the atherosclerosis risk in communities study. *Hypertension* **2013**, HYPERTENSIONAHA. 113.01166.

50. Hao, Y.; Wang, Y.; Xi, L.; Li, G.; Zhao, F.; Qi, Y.; Liu, J.; Zhao, D. A nested case-control study of association between metabolome and hypertension risk. *BioMed Research International* **2016**, *2016*.

51. Aune, D.; Schlesinger, S.; Norat, T.; Riboli, E. Body mass index, abdominal fatness, and the risk of sudden cardiac death: a systematic review and dose–response meta-analysis of prospective studies. Springer: 2018.

52. Collaboration, P.S. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *The Lancet* **2002**, *360*, 1903-1913.

53. Cox, L.M.; Yamanishi, S.; Sohn, J.; Alekseyenko, A.V.; Leung, J.M.; Cho, I.; Kim, S.G.; Li, H.; Gao, Z.; Mahana, D. Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. *Cell* **2014**, *158*, 705-721.

54. Podolsky, S.H. Historical perspective on the rise and fall and rise of antibiotics and human weight gain. *Annals of internal medicine* **2017**, *166*, 133-138.

55. Dahiya, D.K.; Puniya, M.; Shandilya, U.K.; Dhewa, T.; Kumar, N.; Kumar, S.; Puniya, A.K.; Shukla, P. Gut microbiota modulation and its relationship with obesity using prebiotic fibers and probiotics: a review. *Frontiers in microbiology* **2017**, *8*, 563.

56. Park, S.; Ji, Y.; Jung, H.-Y.; Park, H.; Kang, J.; Choi, S.-H.; Shin, H.; Hyun, C.-K.; Kim, K.-T.; Holzapfel, W.H. Lactobacillus plantarum HAC01 regulates gut microbiota and adipose tissue accumulation in a diet-induced obesity murine model. *Applied microbiology and biotechnology* **2017**, *101*, 1605-1614.

57. Lee, S.J.; Bose, S.; Seo, J.-G.; Chung, W.-S.; Lim, C.-Y.; Kim, H. The effects of co-administration of probiotics with herbal medicine on obesity, metabolic endotoxemia and dysbiosis: a randomized double-blind controlled clinical trial. *Clinical nutrition* **2014**, *33*, 973-981.

58. Khalesi, S.; Sun, J.; Buys, N.; Jayasinghe, R. Effect of probiotics on blood pressure: a systematic review and meta-analysis of randomized, controlled trials. *Hypertension* **2014**, *64*, 897-903.

59. Turpin, W.; Espin-Garcia, O.; Xu, W.; Silverberg, M.S.; Kevans, D.; Smith, M.I.; Guttman, D.S.; Griffiths, A.; Panaccione, R.; Otley, A. Association of host genome with intestinal microbial composition in a large healthy cohort. *Nature genetics* **2016**, *48*, 1413.

60. Bonder, M.J.; Kurilshikov, A.; Tigchelaar, E.F.; Mujagic, Z.; Imhann, F.; Vila, A.V.; Deelen, P.; Vatanen, T.; Schirmer, M.; Smeekens, S.P. The effect of host genetics on the gut microbiome. *Nature genetics* **2016**, *48*, 1407.

61. Garrett, M.R.; Rapp, J.P. Defining the blood pressure QTL on chromosome 7 in Dahl rats by a 177-kb congenic segment containing Cyp11b1. *Mammalian Genome* **2003**, *14*, 268-273.

62. Wang, J.; Thingholm, L.B.; Skiecevičienė, J.; Rausch, P.; Kummen, M.; Hov, J.R.; Degenhardt, F.; Heinsen, F.-A.; Rühlemann, M.C.; Szymczak, S. Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nature genetics* **2016**, *48*, 1396.

63. David, L.A.; Maurice, C.F.; Carmody, R.N.; Gootenberg, D.B.; Button, J.E.; Wolfe, B.E.; Ling, A.V.; Devlin, A.S.; Varma, Y.; Fischbach, M.A. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **2014**, *505*, 559.

64. Miranda, P.M.; De Palma, G.; Serkis, V.; Lu, J.; Louis-Auguste, M.P.; McCarville, J.L.; Verdu, E.F.; Collins, S.M.; Bercik, P. High salt diet exacerbates colitis in mice by decreasing Lactobacillus levels and butyrate production. *Microbiome* **2018**, *6*, 57.

65. Wang, C.; Huang, Z.; Yu, K.; Ding, R.; Ye, K.; Dai, C.; Xu, X.; Zhou, G.; Li, C. High-Salt Diet Has a Certain Impact on Protein Digestion and Gut Microbiota: A Sequencing and Proteome Combined Study. *Frontiers in microbiology* **2017**, *8*, 1838.

66. Bier, A.; Braun, T.; Khasbab, R.; Di Segni, A.; Grossman, E.; Haberman, Y.; Leibowitz, A. A high salt diet modulates the gut microbiota and short chain fatty acids production in a salt-sensitive hypertension rat model. *Nutrients* **2018**, *10*, 1154.

67. Derkach, A.; Sampson, J.; Joseph, J.; Playdon, M.C.; Stolzenberg-Solomon, R.Z. Effects of dietary sodium on metabolites: the Dietary Approaches to Stop Hypertension (DASH)–Sodium Feeding Study. *The American journal of clinical nutrition* **2017**, *106*, 1131-1141.

68. Chen, L.; He, F.J.; Dong, Y.; Huang, Y.; Harshfield, G.A.; Zhu, H. Sodium Reduction, Metabolomic Profiling, and Cardiovascular Disease Risk in Untreated Black Hypertensives: A Randomized, Double-Blind, Placebo-Controlled Trial. *Hypertension* **2019**, *74*, 194-200.

69. Lustgarten, M.S.; Price, L.L.; Phillips, E.M.; Kirn, D.R.; Mills, J.; Fielding, R.A. Serum predictors of percent lean mass in young adults. *Journal of strength and conditioning research* **2016**, *30*, 2194-2201.

70. Frame, L.A.; Costa, E.; Jackson, S.A. Current explorations of nutrition and the gut microbiome: a comprehensive evaluation of the review literature. *Nutrition Reviews* **2020**.

71. Zou, J.; Chassaing, B.; Singh, V.; Pellizzon, M.; Ricci, M.; Fythe, M.D.; Kumar, M.V.; Gewirtz, A.T. Fiber-mediated nourishment of gut microbiota protects against diet-induced obesity by restoring IL-22-mediated colonic health. *Cell host & microbe* **2018**, *23*, 41-53. e44.

72. Turnbaugh, P.J.; Ley, R.E.; Mahowald, M.A.; Magrini, V.; Mardis, E.R.; Gordon, J.I. An obesity-associated gut microbiome with increased capacity for energy harvest. *nature* **2006**, *444*, 1027.

73. Goffredo, M.; Mass, K.; Parks, E.J.; Wagner, D.A.; McClure, E.A.; Graf, J.; Savoye, M.; Pierpont, B.; Cline, G.; Santoro, N. Role of gut microbiota and short chain fatty acids in modulating energy harvest and fat partitioning in youth. *The Journal of Clinical Endocrinology & Metabolism* **2016**, *101*, 4367-4376.

74. Samuel, B.S.; Shaito, A.; Motoike, T.; Rey, F.E.; Backhed, F.; Manchester, J.K.; Hammer, R.E.; Williams, S.C.; Crowley, J.; Yanagisawa, M. Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41. *Proceedings of the National Academy of Sciences* **2008**, *105*, 16767-16772.

75. De Filippis, F.; Pellegrini, N.; Vannini, L.; Jeffery, I.B.; La Storia, A.; Laghi, L.; Serrazanetti, D.I.; Di Cagno, R.; Ferrocino, I.; Lazzi, C. High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. *Gut* **2016**, *65*, 1812-1821.

76. Mueller, N.T.; Zhang, M.; Juraschek, S.P.; Miller, E.R.; Appel, L.J. Effects of high-fiber diets enriched with carbohydrate, protein, or unsaturated fat on circulating short chain fatty acids: results from the OmniHeart randomized trial. *The American Journal of Clinical Nutrition* **2020**.

77. Jeffery, I.B.; O'Toole, P.W. Diet-microbiota interactions and their implications for healthy living. *Nutrients* **2013**, *5*, 234-252.

78. la Cuesta-Zuluaga, D.; Mueller, N.T.; Álvarez-Quintero, R.; Velásquez-Mejía, E.P.; Sierra, J.A.; Corrales-Agudelo, V.; Carmona, J.A.; Abad, J.M.; Escobar, J.S. Higher fecal short-chain fatty acid levels are associated with gut microbiome dysbiosis, obesity, hypertension and cardiometabolic disease risk factors. *Nutrients* **2019**, *11*, 51.

79. Natarajan, N.; Pluznick, J.L. Olfaction in the kidney:'smelling'gut microbial metabolites. *Experimental physiology* **2016**, *101*, 478-481.

80. Pluznick, J. A novel SCFA receptor, the microbiota, and blood pressure regulation. *Gut microbes* **2014**, *5*, 202-207.

81. LeBlanc, J.G.; Chain, F.; Martín, R.; Bermúdez-Humarán, L.G.; Courau, S.; Langella, P. Beneficial effects on host energy metabolism of short-chain fatty acids and vitamins produced by commensal and probiotic bacteria. *Microbial cell factories* **2017**, *16*, 79.

82. Gao, Z.; Yin, J.; Zhang, J.; Ward, R.E.; Martin, R.J.; Lefevre, M.; Cefalu, W.T.; Ye, J. Butyrate improves insulin sensitivity and increases energy expenditure in mice. *Diabetes* **2009**, *58*, 1509-1517.

83. Chambers, E.S.; Viardot, A.; Psichas, A.; Morrison, D.J.; Murphy, K.G.; Zac-Varghese, S.E.; MacDougall, K.; Preston, T.; Tedford, C.; Finlayson, G.S. Effects of targeted delivery of propionate to the human colon on appetite regulation, body weight maintenance and adiposity in overweight adults. *Gut* **2015**, *64*, 1744-1754.

84. Byrne, C.; Chambers, E.; Morrison, D.; Frost, G. The role of short chain fatty acids in appetite regulation and energy homeostasis. *International journal of obesity* **2015**, *39*, 1331-1338.

85. Chambers, E.S.; Byrne, C.S.; Aspey, K.; Chen, Y.; Khan, S.; Morrison, D.J.; Frost, G. Acute oral sodium propionate supplementation raises resting energy expenditure and lipid oxidation in fasted humans. *Diabetes, Obesity and Metabolism* **2018**, *20*, 1034-1039.

86. Brahe, L.K.; Astrup, A.; Larsen, L.H. Is butyrate the link between diet, intestinal microbiota and obesity-related metabolic diseases? *Obesity reviews* **2013**, *14*, 950-959.

87. Isken, F.; Klaus, S.; Osterhoff, M.; Pfeiffer, A.F.; Weickert, M.O. Effects of long-term soluble vs. insoluble dietary fiber intake on high-fat diet-induced obesity in C57BL/6J mice. *The Journal of nutritional biochemistry* **2010**, *21*, 278-284.

88. Kim, K.N.; Yao, Y.; Ju, S.Y. Short chain fatty acids and fecal microbiota abundance in humans with obesity: A systematic review and meta-analysis. *Nutrients* **2019**, *11*, 2512.

89. Schwiertz, A.; Taras, D.; Schäfer, K.; Beijer, S.; Bos, N.A.; Donus, C.; Hardt, P.D. Microbiota and SCFA in lean and overweight healthy subjects. *Obesity* **2010**, *18*, 190-195.

90. Rahat-Rozenbloom, S.; Fernandes, J.; Gloor, G.B.; Wolever, T.M. Evidence for greater production of colonic short-chain fatty acids in overweight than lean humans. *International journal of obesity* **2014**, *38*, 1525-1531.

91. Miranda, V.P.N.; dos Santos Amorim, P.R.; Bastos, R.R.; de Faria, E.R.; de Castro Moreira, M.E.; do Carmo Castro Franceschini, S.; do Carmo Gouveia Peluzio, M.; de Luces Fortes Ferreira, C.L.; Priore, S.E. Abundance of Gut Microbiota, Concentration of Short-Chain Fatty Acids, and Inflammatory Markers Associated with Elevated Body Fat, Overweight, and Obesity in Female Adolescents. *Mediators of Inflammation* **2019**, *2019*.

92. Brooks, A.W.; Priya, S.; Blekhman, R.; Bordenstein, S.R. Gut microbiota diversity across ethnicities in the United States. *PLoS biology* **2018**, *16*.

93. Meyer, K.A.; Bennett, B.J. Diet and gut microbial function in metabolic and cardiovascular disease risk. *Current diabetes reports* **2016**, *16*, 93.

94. Firestone, M.J.; Beasley, J.M.; Kwon, S.C.; Ahn, J.; Trinh-Shevrin, C.; Yi, S.S. Asian American dietary sources of sodium and salt behaviors compared with other racial/ethnic groups, NHANES, 2011-2012. *Ethnicity & disease* **2017**, *27*, 241.

95. Du, S.; Neiman, A.; Batis, C.; Wang, H.; Zhang, B.; Zhang, J.; Popkin, B.M. Understanding the patterns and trends of sodium intake, potassium intake, and sodium to potassium ratio and their effect on hypertension in China–. *The American journal of clinical nutrition* **2013**, *99*, 334-343.

96. Hipgrave, D.B.; Chang, S.; Li, X.; Wu, Y. Salt and sodium intake in China. *Jama* **2016**, *315*, 703-705.

97. Gordon-Larsen, P.; Attard, S.M.; Howard, A.G.; Popkin, B.M.; Zhang, B.; Du, S.; Guilkey, D.K. Accounting for Selectivity Bias and Correlation Across the Sequence From Elevated Blood Pressure to Hypertension Diagnosis and Treatment. *American journal of hypertension* **2017**, *31*, 63-71.

98. Popkin, B.M.; Du, S.; Zhai, F.; Zhang, B. Cohort Profile: The China Health and Nutrition Survey—monitoring and understanding socio-economic and health change in China, 1989–2011. *International journal of epidemiology* **2009**, *39*, 1435-1440.

99. Wang, L. Report of China nationwide nutrition and health survey 2002 (1): summary report. *Beijing: People's Medical Publishing House* **2005**, 18-45.

100. Yao, M.; Lichtenstein, A.; Roberts, S.; Ma, G.; Gao, S.; Tucker, K.; McCrory, M. Relative influence of diet and physical activity on cardiovascular risk factors in urban Chinese adults. *International journal of obesity* **2003**, *27*, 920.

101. Caporaso, J.G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F.D.; Costello, E.K.; Fierer, N.; Pena, A.G.; Goodrich, J.K.; Gordon, J.I. QIIME allows analysis of high-throughput community sequencing data. *Nature methods* **2010**, *7*, 335.

102. Haas, B.J.; Gevers, D.; Earl, A.M.; Feldgarden, M.; Ward, D.V.; Giannoukos, G.; Ciulla, D.; Tabbaa, D.; Highlander, S.K.; Sodergren, E. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome research* **2011**, *21*, 494-504.

103. Truong, D.T.; Franzosa, E.A.; Tickle, T.L.; Scholz, M.; Weingart, G.; Pasolli, E.; Tett, A.; Huttenhower, C.; Segata, N. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature methods* **2015**, *12*, 902-903.

104. Jones, R.B.; Zhu, X.; Moan, E.; Murff, H.J.; Ness, R.M.; Seidner, D.L.; Sun, S.; Yu, C.; Dai, Q.; Fodor, A.A. Inter-niche and inter-individual variation in gut microbial community assessment using stool, rectal swab, and mucosal samples. *Scientific reports* **2018**, *8*, 1-12.

105. Evans, A.M.; DeHaven, C.D.; Barrett, T.; Mitchell, M.; Milgram, E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Analytical chemistry* **2009**, *81*, 6656-6667.

106. Long, T.; Hicks, M.; Yu, H.-C.; Biggs, W.H.; Kirkness, E.F.; Menni, C.; Zierer, J.; Small, K.S.; Mangino, M.; Messier, H. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nature genetics* **2017**, *49*, 568.

107. Zhou, B.-F. Predictive values of body mass index and waist circumference for risk factors of certain related diseases in Chinese adults--study on optimal cut-off points of body mass index and waist circumference in Chinese adults. *Biomedical and environmental sciences: BES* **2002**, *15*, 83-96.

108. Peng, Y.; Li, W.; Wang, Y.; Bo, J.; Chen, H. The cut-off point and boundary values of waist-to-height ratio as an indicator for cardiovascular risk factors in Chinese adults from the PURE study. *PLoS One* **2015**, *10*, e0144539.

109. Jones-Smith, J.C.; Popkin, B.M. Understanding community context and adult health changes in China: development of an urbanicity scale. *Social science & medicine* **2010**, *71*, 1436-1446.

110. Cook, N.R.; Appel, L.J.; Whelton, P.K. Lower levels of sodium intake and reduced cardiovascular risk. *Circulation* **2014**, *129*, 981-989.

111. He, F.J.; MacGregor, G.A. Salt reduction lowers cardiovascular risk: meta-analysis of outcome trials. *The Lancet* **2011**, *378*, 380-382.

112. Elijovich, F.; Weinberger, M.H.; Anderson, C.A.; Appel, L.J.; Bursztyn, M.; Cook, N.R.; Dart, R.A.; Newton-Cheh, C.H.; Sacks, F.M.; Laffer, C.L. Salt sensitivity of blood pressure: a scientific statement from the American Heart Association. *Hypertension* **2016**, *68*, e7-e46.

113. Kong, Y.W.; Baqar, S.; Jerums, G.; Ekinci, E.I. Sodium and its role in cardiovascular disease–the debate continues. *Frontiers in endocrinology* **2016**, *7*, 164.

114. Bartolomaeus, H.; Balogh, A.; Yakoub, M.; Homann, S.; Markó, L.; Höges, S.; Tsvetkov, D.; Krannich, A.; Wundersitz, S.; Avery, E.G. Short-chain fatty acid propionate protects from hypertensive cardiovascular damage. *Circulation* **2019**, *139*, 1407-1421.

115. Härtl, G. WHO issues new guidance on dietary

salt and potassium. . Availabe online: https://www.who.int/mediacentre/news/notes/2013/salt_potassium_20130131/en/ (accessed on November).

116. Oksanen, J.; Kindt, R.; Legendre, P.; O'Hara, B.; Stevens, M.; Oksanen, M.; Suggests, M. The vegan package: community ecology package. *R package version* **2007**, *1*, 15-11.

117. McArdle, B.H.; Anderson, M.J. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* **2001**, *82*, 290-297.

118. Sun, S.; Wang, H.; Tsilimigras, M.C.; Howard, A.G.; Sha, W.; Du, S.; Sioda, M.; Fodor, A.; Gordon-Larsen, P. Geographic variation profoundly impacts the reproducibility of the associations between host factors and human gut microbiota in China. *Submitted* **2020**.

119. Winglee, K.; Howard, A.G.; Sha, W.; Gharaibeh, R.Z.; Liu, J.; Jin, D.; Fodor, A.A.; Gordon-Larsen, P. Recent urbanization in China is correlated with a Westernized microbiome encoding increased virulence and antibiotic resistance genes. *Microbiome* **2017**, *5*, 121.

120. Popkin, B.M.; Du, S. Dynamics of the nutrition transition toward the animal foods sector in China and its implications: a worried perspective. *The Journal of nutrition* **2003**, *133*, 3898S-3906S.

121. Ding, M.; Zeleznik, O.A.; Guasch-Ferre, M.; Hu, J.; Lasky-Su, J.; Lee, I.-M.; Jackson, R.D.; Shadyab, A.H.; LaMonte, M.J.; Clish, C. Metabolome-wide Association Study with Habitual Physical Activity in Four Prospective Cohort Studies. *American journal of epidemiology* **2019**.

122. Lee, S.H.; Yun, Y.; Kim, S.J.; Lee, E.-J.; Chang, Y.; Ryu, S.; Shin, H.; Kim, H.-L.; Kim, H.-N.; Lee, J.H. Association between cigarette smoking status and composition of gut microbiota: population-based cross-sectional study. *Journal of clinical medicine* **2018**, *7*, 282.

123. Imhann, F.; Bonder, M.J.; Vila, A.V.; Fu, J.; Mujagic, Z.; Vork, L.; Tigchelaar, E.F.; Jankipersadsing, S.A.; Cenit, M.C.; Harmsen, H.J. Proton pump inhibitors affect the gut microbiome. *Gut* **2016**, *65*, 740-748.

124. Manning, A.K.; LaValley, M.; Liu, C.T.; Rice, K.; An, P.; Liu, Y.; Miljkovic, I.; Rasmussen-Torvik, L.; Harris, T.B.; Province, M.A. Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP× environment regression coefficients. *Genetic epidemiology* **2011**, *35*, 11-18.

125. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *Journal of machine learning research* **2011**, *12*, 2825-2830.

126. Dietterich, T.G. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation* **1998**, *10*, 1895-1923.

127. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **1995**, *57*, 289-300.

128. Lowy, F.D. Staphylococcus aureus infections. *New England journal of medicine* **1998**, *339*, 520-532.

129. Chuang, C.-H.; Wang, Y.-H.; Chang, H.-J.; Chen, H.-L.; Huang, Y.-C.; Lin, T.-Y.; Ozer, E.A.; Allen, J.P.; Hauser, A.R.; Chiu, C.-H. Shanghai fever: a distinct Pseudomonas aeruginosa enteric disease. *Gut* **2014**, *63*, 736-743.

130. Ottosson, F.; Brunkwall, L.; Ericson, U.; Nilsson, P.M.; Almgren, P.; Fernandez, C.; Melander, O.; Orho-Melander, M. Connection between BMI-related plasma metabolite profile and gut microbiota. *The Journal of Clinical Endocrinology & Metabolism* **2018**, *103*, 1491-1501.

131. Menni, C.; Jackson, M.A.; Pallister, T.; Steves, C.J.; Spector, T.D.; Valdes, A.M. Gut microbiome diversity and high-fibre intake are related to lower long-term weight gain. *International Journal of Obesity* **2017**, *41*, 1099.

132. Lippert, K.; Kedenko, L.; Antonielli, L.; Kedenko, I.; Gemeier, C.; Leitner, M.; Kautzky-Willer, A.; Paulweber, B.; Hackl, E. Gut microbiota dysbiosis associated with glucose metabolism disorders and the metabolic syndrome in older adults. *Beneficial microbes* **2017**, *8*, 545-556.

133. Larrosa, M.; Luceri, C.; Vivoli, E.; Pagliuca, C.; Lodovici, M.; Moneti, G.; Dolara, P. Polyphenol metabolites from colonic microbiota exert anti-inflammatory activity on different inflammation models. *Mol Nutr Food Res* **2009**, *53*, 1044-1054, doi:10.1002/mnfr.200800446.

134. Chambers, E.S.; Preston, T.; Frost, G.; Morrison, D.J. Role of gut Microbiota-Generated short-chain fatty acids in metabolic and cardiovascular health. *Current nutrition reports* **2018**, *7*, 198-206.

135. Kim, S.; Thapa, I.; Zhang, L.; Ali, H. A novel graph theoretical approach for modeling microbiomes and inferring microbial ecological relationships. *BMC genomics* **2019**, *20*, 1-13.

136. Monteleone, I.; Marafini, I.; Dinallo, V.; Di Fusco, D.; Troncone, E.; Zorzi, F.; Laudisi, F.; Monteleone, G. Sodium chloride–enriched diet enhanced inflammatory cytokine production and exacerbated experimental colitis in mice. *Journal of Crohn's and Colitis* **2017**, *11*, 237-245.

137. Clavel, T.; Mapesa, J.O. Phenolics in human nutrition: importance of the intestinal microbiome for isoflavone and lignan bioavailability. *Natural products. Amsterdam: Elsevier* **2013**, 2433-2463.

138. Wen, X.; Zhou, L.; Stamler, J.; Chan, Q.; Van Horn, L.; Daviglus, M.L.; Dyer, A.R.; Elliott, P.; Ueshima, H.; Miura, K. Agreement between 24-h dietary recalls and 24-h urine collections for estimating sodium intake in China, Japan, UK, USA: the International Study of Macro-and Micro-nutrients and Blood Pressure. *Journal of hypertension* **2019**, *37*, 814.

139. Zheng, Y.; Yu, B.; Alexander, D.; Couper, D.J.; Boerwinkle, E. Medium-term variability of the human serum metabolome in the Atherosclerosis Risk in Communities (ARIC) study. *Omics: a journal of integrative biology* **2014**, *18*, 364-373.

140. Fryar, C.D.; Ostchega, Y.; Hales, C.M.; Zhang, G.; Kruszon-Moran, D. Hypertension prevalence and control among adults: United States, 2015-2016. **2017**.

141. Lu, X.; Wang, L.; Lin, X.; Huang, J.; Charles Gu, C.; He, M.; Shen, H.; He, J.; Zhu, J.; Li, H. Genome-wide association study in Chinese identifies novel loci for blood pressure and hypertension. *Human molecular genetics* **2014**, *24*, 865-874.

142. Rothschild, D.; Weissbrod, O.; Barkan, E.; Kurilshikov, A.; Korem, T.; Zeevi, D.; Costea, P.I.; Godneva, A.; Kalka, I.N.; Bar, N. Environment dominates over host genetics in shaping human gut microbiota. *Nature* **2018**, *555*, 210-215.

143. Yang, T.; Richards, E.M.; Pepine, C.J.; Raizada, M.K. The gut microbiota and the brain–gut–kidney axis in hypertension and chronic kidney disease. *Nature Reviews Nephrology* **2018**, *14*, 442-456.

144. Menni, C.; Graham, D.; Kastenmüller, G.; Alharbi, N.H.; Alsanosi, S.M.; McBride, M.; Mangino, M.; Titcombe, P.; Shin, S.-Y.; Psatha, M. Metabolomic identification of a novel pathway of blood pressure regulation involving hexadecanedioate. *Hypertension* **2015**, *66*, 422-429.

145. Silveira-Nunes, G.; Durso, D.F.; Cunha, E.H.M.; Maioli, T.U.; Vieira, A.T.; Speziali, E.; Corrêa-Oliveira, R.; Martins-Filho, O.A.; Teixeira-Carvalho, A.; Franceschi, C. Hypertension Is Associated With Intestinal Microbiota Dysbiosis and Inflammation in a Brazilian Population. *Frontiers in pharmacology* **2020**, *11*, 258.

146. Xi, B.; Liang, Y.; Reilly, K.H.; Wang, Q.; Hu, Y.; Tang, W. Trends in prevalence, awareness, treatment, and control of hypertension among Chinese adults 1991–2009. *International journal of cardiology* **2012**, *158*, 326-329.

147. Levey, A.S.; Stevens, L.A.; Schmid, C.H.; Zhang, Y.L.; Castro, A.F.; Feldman, H.I.; Kusek, J.W.; Eggers, P.; Van Lente, F.; Greene, T. A new equation to estimate glomerular filtration rate. *Annals of internal medicine* **2009**, *150*, 604-612.

148. Peet, R.K. The measurement of species diversity. *Annual review of ecology and systematics* **1974**, *5*, 285-307.

149. Faith, D.P.; Minchin, P.R.; Belbin, L. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* **1987**, *69*, 57-68.

150. Anderson, M.J. Permutational multivariate analysis of variance (PERMANOVA). *Wiley StatsRef: Statistics Reference Online* **2014**, 1-15.

151. Sun, S.; Wang, H.; Tsilimigras, M.C.; Howard, A.G.; Sha, W.; Du, S.; Sioda, M.; Fodor, A.; Gordon-Larsen, P. Geographic variation profoundly impacts the reproducibility of the associations between host factors and human gut microbiota in China. *Manuscript in preparation* **2019**.

152. Wang, Y.; Wang, H.; Howard, A.G.; Tsilimigras, M.C.; Avery, C.L.; Meyer, K.A.; Sha, W.; Sun, S.; Zhang, J.; Su, C., et al. Associations of sodium and potassium consumption with gut microbiota and host metabolites in a population-based study of Chinese adults. *Submitted* **2020**.

153. Sekula, P.; Goek, O.-N.; Quaye, L.; Barrios, C.; Levey, A.S.; Römisch-Margl, W.; Menni, C.; Yet, I.; Gieger, C.; Inker, L.A. A metabolome-wide association study of kidney function and disease in the general population. *Journal of the American Society of Nephrology* **2016**, *27*, 1175-1188.

154. Antonelli, J.; Claggett, B.L.; Henglin, M.; Kim, A.; Ovsak, G.; Kim, N.; Deng, K.; Rao, K.; Tyagi, O.; Watrous, J.D. Statistical workflow for feature selection in human metabolomics data. *Metabolites* **2019**, *9*, 143.

155. Conway, J.M.; Huffcutt, A.I. A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational research methods* **2003**, *6*, 147-168.

156. McDonald, J.H. *Handbook of biological statistics*; sparky house publishing Baltimore, MD: 2009; Vol. 2.

157. Holmes, E.; Loo, R.L.; Stamler, J.; Bictash, M.; Yap, I.K.; Chan, Q.; Ebbels, T.; De Iorio, M.; Brown, I.J.; Veselkov, K.A. Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* **2008**, *453*, 396.

158. Chen, Y.-Y.; Chen, D.-Q.; Chen, L.; Liu, J.-R.; Vaziri, N.D.; Guo, Y.; Zhao, Y.-Y. Microbiome–metabolome reveals the contribution of gut–kidney axis on kidney disease. *Journal of translational medicine* **2019**, *17*, 1-11.

159. Pryde, S.E.; Duncan, S.H.; Hold, G.L.; Stewart, C.S.; Flint, H.J. The microbiology of butyrate formation in the human colon. *FEMS microbiology letters* **2002**, *217*, 133-139.

160. Mels, C.M.; Schutte, A.E.; Erasmus, E.; Huisman, H.W.; Schutte, R.; Fourie, C.M.; Kruger, R.; Van Rooyen, J.M.; Smith, W.; Malan, N.T. L-carnitine and long-chain acylcarnitines are positively correlated with ambulatory blood pressure in humans: the SABPA study. *Lipids* **2013**, *48*, 63-73.

161. Adams, S.H.; Hoppel, C.L.; Lok, K.H.; Zhao, L.; Wong, S.W.; Minkler, P.E.; Hwang, D.H.; Newman, J.W.; Garvey, W.T. Plasma acylcarnitine profiles suggest incomplete long-chain fatty acid β-oxidation and altered tricarboxylic acid cycle activity in type 2 diabetic African-American women. *The Journal of nutrition* **2009**, *139*, 1073-1081.

162. Harris, W.S.; Shearer, G.C. Omega-6 fatty acids and cardiovascular disease: friend, not foe? Am Heart Assoc: 2014.

163. Chaurasia, B.; Summers, S.A. Ceramides–lipotoxic inducers of metabolic disorders. *Trends in Endocrinology & Metabolism* **2015**, *26*, 538-550.

164. Summers, S.A.; Chaurasia, B.; Holland, W.L. Metabolic Messengers: ceramides. *Nature Metabolism* **2019**, 1-8.

165. Havulinna, A.S.; Sysi-Aho, M.; Hilvo, M.; Kauhanen, D.; Hurme, R.; Ekroos, K.; Salomaa, V.; Laaksonen, R. Circulating ceramides predict cardiovascular outcomes in the population-based FINRISK 2002 cohort. *Arteriosclerosis, thrombosis, and vascular biology* **2016**, *36*, 2424-2430.

166. Laaksonen, R.; Ekroos, K.; Sysi-Aho, M.; Hilvo, M.; Vihervaara, T.; Kauhanen, D.; Suoniemi, M.; Hurme, R.; März, W.; Scharnagl, H. Plasma ceramides predict cardiovascular death in patients with stable coronary artery disease and acute coronary syndromes beyond LDL-cholesterol. *European heart journal* **2016**, *37*, 1967-1976.

167. Poss, A.M.; Maschek, J.A.; Cox, J.E.; Hauner, B.J.; Hopkins, P.N.; Hunt, S.C.; Holland, W.L.; Summers, S.A.; Playdon, M.C. Machine learning reveals serum sphingolipids as cholesterol-independent biomarkers of coronary artery disease. *The Journal of Clinical Investigation* **2020**, *130*.

168. Yeh, T.-L.; Chen, H.-H.; Tsai, S.-Y.; Lin, C.-Y.; Liu, S.-J.; Chien, K.-L. The Relationship between Metabolically Healthy Obesity and the Risk of Cardiovascular Disease: A Systematic Review and Meta-Analysis. *Journal of clinical medicine* **2019**, *8*, 1228.

169. Ashwell, M.; Gibson, S. Waist-to-height ratio as an indicator of 'early health risk': simpler and more predictive than using a 'matrix'based on BMI and waist circumference. *BMJ open* **2016**, *6*, e010159.

170. Hu, X.; Gao, J.; Zhang, Q.; Fu, Y.; Li, K.; Zhu, S.; Li, D. Soy fiber improves weight loss and lipid profile in overweight and obese adults: a randomized controlled trial. *Molecular nutrition & food research* **2013**, *57*, 2147-2154.

171. Esposito, K.; Kastorini, C.-M.; Panagiotakos, D.B.; Giugliano, D. Mediterranean diet and weight loss: meta-analysis of randomized controlled trials. *Metabolic syndrome and related disorders* **2011**, *9*, 1-12.

172. Inoue, Y.; Howard, A.G.; Thompson, A.L.; Gordon-Larsen, P. Secular change in the association between urbanisation and abdominal adiposity in China (1993–2011). *J Epidemiol Community Health* **2018**, *72*, 484-490.

173. Yan, R.; Li, W.; Yin, L.; Wang, Y.; Bo, J.; Investigators, P.C.; Investigators, P.C.; Liu, L.; Liu, B.; Hu, B. Cardiovascular diseases and risk-factor burden in urban and rural communities in high-, middle-, and low-income regions of china: a large community-based epidemiological study. *Journal of the American Heart Association* **2017**, *6*, e004445.

174. Vogt, J.A.; Wolever, T.M. Fecal acetate is inversely related to acetate absorption from the human rectum and distal colon. *The Journal of nutrition* **2003**, *133*, 3145-3148.

175. Oliphant, K.; Allen-Vercoe, E. Macronutrient metabolism by the human gut microbiome: major fermentation by-products and their impact on host health. *Microbiome* **2019**, *7*, 91.

176. Heimann, E.; Nyman, M.; Pålbrink, A.-K.; Lindkvist-Petersson, K.; Degerman, E. Branched short-chain fatty acids modulate glucose and lipid metabolism in primary adipocytes. *Adipocyte* **2016**, *5*, 359-368.

177. Boets, E.; Deroover, L.; Houben, E.; Vermeulen, K.; Gomand, S.V.; Delcour, J.A.; Verbeke, K. Quantification of in vivo colonic short chain fatty acid production from inulin. *Nutrients* **2015**, *7*, 8916-8929.

178. Layden, B.T.; Yalamanchi, S.K.; Wolever, T.M.; Dunaif, A.; Lowe Jr, W.L. Negative association of acetate with visceral adipose tissue and insulin levels. *Diabetes, metabolic syndrome and obesity: targets and therapy* **2012**, *5*, 49.

179. Yang, J.; Keshavarzian, A.; Rose, D.J. Impact of dietary fiber fermentation from cereal grains on metabolite production by the fecal microbiota from normal weight and obese individuals. *Journal of medicinal food* **2013**, *16*, 862-867.

180. Salazar, N.; Dewulf, E.M.; Neyrinck, A.M.; Bindels, L.B.; Cani, P.D.; Mahillon, J.; de Vos, W.M.; Thissen, J.-P.; Gueimonde, M.; Clara, G. Inulin-type fructans modulate intestinal Bifidobacterium species populations and decrease fecal short-chain fatty acids in obese women. *Clinical nutrition* **2015**, *34*, 501-507.

181. Baxter, N.T.; Schmidt, A.W.; Venkataraman, A.; Kim, K.S.; Waldron, C.; Schmidt, T.M. Dynamics of human gut microbiota and short-chain fatty acids in response to dietary interventions with three fermentable fibers. *MBio* **2019**, *10*, e02566-02518.

182. Ríos-Covián, D.; Ruas-Madiedo, P.; Margolles, A.; Gueimonde, M.; de los Reyes-Gavilán, C.G.; Salazar, N. Intestinal short chain fatty acids and their link with diet and human health. *Frontiers in microbiology* **2016**, *7*, 185.

183. Cummings, J.; Pomare, E.; Branch, W.; Naylor, C.; Macfarlane, G. Short chain fatty acids in human large intestine, portal, hepatic and venous blood. *Gut* **1987**, *28*, 1221-1227.

184. Barcenilla, A.; Pryde, S.E.; Martin, J.C.; Duncan, S.H.; Stewart, C.S.; Henderson, C.; Flint, H.J. Phylogenetic relationships of butyrate-producing bacteria from the human gut. *Appl. Environ. Microbiol.* **2000**, *66*, 1654-1661.

185. Louis, P.; Flint, H.J. Formation of propionate and butyrate by the human colonic microbiota. *Environmental microbiology* **2017**, *19*, 29-41.

186. Reichardt, N.; Duncan, S.H.; Young, P.; Belenguer, A.; Leitch, C.M.; Scott, K.P.; Flint, H.J.; Louis, P. Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *The ISME journal* **2014**, *8*, 1323.

187. Vital, M.; Howe, A.C.; Tiedje, J.M. Revealing the bacterial butyrate synthesis pathways by analyzing (meta) genomic data. *MBio* **2014**, *5*, e00889-00814.

188. Louis, P.; Young, P.; Holtrop, G.; Flint, H.J. Diversity of human colonic butyrate-producing bacteria revealed by analysis of the butyryl-CoA: acetate CoA-transferase gene. *Environmental microbiology* **2010**, *12*, 304-314.

189. Muñoz-Tamayo, R.; Laroche, B.; Walter, É.; Doré, J.; Duncan, S.H.; Flint, H.J.; Leclerc, M. Kinetic modelling of lactate utilization and butyrate production by key human colonic bacterial species. *FEMS microbiology ecology* **2011**, *76*, 615-624.

190. Duncan, S.H.; Louis, P.; Flint, H.J. Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. *Appl. Environ. Microbiol.* **2004**, *70*, 5810-5817.

191. Belenguer, A.; Duncan, S.H.; Calder, A.G.; Holtrop, G.; Louis, P.; Lobley, G.E.; Flint, H.J. Two routes of metabolic cross-feeding between Bifidobacterium adolescentis and butyrate-producing anaerobes from the human gut. *Appl. Environ. Microbiol.* **2006**, *72*, 3593-3599.

192. Patterson, A.M.; Mulder, I.E.; Travis, A.J.; Lan, A.; Cerf-Bensussan, N.; Gaboriau-Routhiau, V.; Garden, K.; Logan, E.; Delday, M.I.; Coutts, A.G. Human gut symbiont Roseburia hominis promotes and regulates innate immunity. *Frontiers in immunology* **2017**, *8*, 1166.

193. Duncan, S.H.; Holtrop, G.; Lobley, G.E.; Calder, A.G.; Stewart, C.S.; Flint, H.J. Contribution of acetate to butyrate formation by human faecal bacteria. *British Journal of Nutrition* **2004**, *91*, 915-923.

194. Crost, E.H.; Tailford, L.E.; Le Gall, G.; Fons, M.; Henrissat, B.; Juge, N. Utilisation of mucin glycans by the human gut symbiont Ruminococcus gnavus is strain-dependent. *PloS one* **2013**, *8*, e76341.

195. Wrzosek, L.; Miquel, S.; Noordine, M.-L.; Bouet, S.; Chevalier-Curt, M.J.; Robert, V.; Philippe, C.; Bridonneau, C.; Cherbuy, C.; Robbe-Masselot, C. Bacteroides thetaiotaomicron and

Faecalibacterium prausnitzii influence the production of mucus glycans and the development of goblet cells in the colonic epithelium of a gnotobiotic model rodent. *BMC biology* **2013**, *11*, 61.

196. Laverde Gomez, J.A.; Mukhopadhya, I.; Duncan, S.H.; Louis, P.; Shaw, S.; Collie-Duguid, E.; Crost, E.; Juge, N.; Flint, H.J. Formate cross-feeding and cooperative metabolic interactions revealed by transcriptomics in co-cultures of acetogenic and amylolytic human colonic bacteria. *Environmental microbiology* **2019**, *21*, 259-271.

197. Sakon, H.; Nagai, F.; Morotomi, M.; Tanaka, R. Sutterella parvirubra sp. nov. and Megamonas funiformis sp. nov., isolated from human faeces. *International journal of systematic and evolutionary microbiology* **2008**, *58*, 970-975.

198. Kelly, W.J.; Henderson, G.; Pacheco, D.M.; Li, D.; Reilly, K.; Naylor, G.E.; Janssen, P.H.; Attwood, G.T.; Altermann, E.; Leahy, S.C. The complete genome sequence of Eubacterium limosum SA11, a metabolically versatile rumen acetogen. *Standards in genomic sciences* **2016**, *11*, 26.

199. Durand, M.; Bernalier, A.; Cummings, J.; Rombeau, J.; Sakata, T. Reductive acetogenesis in animal and human gut. *Physiological and clinical aspects of shortchain fatty acids* **1993**, 107-117.

200. Sato, T.; Matsumoto, K.; Okumura, T.; Yokoi, W.; Naito, E.; Yoshida, Y.; Nomoto, K.; Ito, M.; Sawada, H. Isolation of lactate-utilizing butyrate-producing bacteria from human feces and in vivo administration of Anaerostipes caccae strain L2 and galacto-oligosaccharides in a rat model. *FEMS microbiology ecology* **2008**, *66*, 528-536.

201. Sivieri, K.; Morales, M.L.V.; Adorno, M.A.T.; Sakamoto, I.K.; Saad, S.M.I.; Rossi, E.A. Lactobacillus acidophilus CRL 1014 improved "gut health" in the SHIME® reactor. *BMC gastroenterology* **2013**, *13*, 100.

202. Derrien, M.; Vaughan, E.E.; Plugge, C.M.; de Vos, W.M. Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *International journal of systematic and evolutionary microbiology* **2004**, *54*, 1469-1476.

203. Davey Smith, G.; Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human molecular genetics* **2014**, *23*, R89-R98.

204. Cobb, L.K.; Anderson, C.A.; Elliott, P.; Hu, F.B.; Liu, K.; Neaton, J.D.; Whelton, P.K.; Woodward, M.; Appel, L.J. Methodological issues in cohort studies that relate sodium intake to cardiovascular disease outcomes: a science advisory from the American Heart Association. *Circulation* **2014**, *129*, 1173-1186.

205. Medina-Remón, A.; Tresserra-Rimbau, A.; Pons, A.; Tur, J.A.; Martorell, M.; Ros, E.; Buil-Cosiales, P.; Sacanella, E.; Covas, M.I.; Corella, D. Effects of total dietary polyphenols on plasma nitric oxide and blood pressure in a high cardiovascular risk cohort. The PREDIMED randomized trial. *Nutrition, Metabolism and Cardiovascular Diseases* **2015**, *25*, 60-67.

206. van Goor, H.; van den Born, J.C.; Hillebrands, J.-L.; Joles, J.A. Hydrogen sulfide in hypertension. *Current opinion in nephrology and hypertension* **2016**, *25*, 107-113.