보건학박사 학위논문

The effects of human genetic and environmental factors on the gut and airway microbiome

인체의 유전적 요인과 환경적 요인이 장내와 호흡기 마이크로비옴에 미치는 영향

2015년 8월

서울대학교 대학원

보건학과 보건미생물학 전공

임 미 영

The effects of human genetic and environmental factors on the gut and airway microbiome

Mi Young Lim

Department of Public Health,
Graduate School of Public Health,
Seoul National University

A thesis submitted to the faculty of the Seoul National University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Public Health, Graduate School of Public Health.

Seoul, Korea
2015.06

Sungkyoon Kim

Domyung Paek

Hai Won Chung

Mina Rho

GwangPyo Ko

인체의 유전적 요인과 환경적 요인이 장내와 호흡기 마이크로비옴에 미치는 영향

지도교수 고 광 표

위 논문을 임미영 박사학위 논문으로 제출함.

2015년 3월

서울대학교 보건대학원

보건학과

임 미 영

임미영의 박사학위 논문을 인준함.

2015년 6월

위 원 장 김 성 균 (인)

부위원장 백 도 명 (인)

위 원 정 해 원 (인)

위 원 노 미 나 (인)

위 원 고 광 표 (인)

ABSTRACT

The effects of human genetic and environmental factors on the gut and airway microbiome

Mi Young Lim
Department of Public Health
Graduate School of Public Health
Seoul National University

The human microbiome consists of an enormous variety of microorganisms including bacteria, archaea, fungi, and viruses that reside in all body surfaces that are exposed to the external environment such as the skin, oral cavity, airway, vagina, and gut. Although the human microbiome has long been known to play important roles in human health and disease, only the recent advances in high-throughput sequencing methods and analytic techniques have enable detailed investigation of the human microbiome composition, function, and host-microbe interactions. Through these studies, it has become apparent that in the healthy state, the human microbiota is involved in a broad range of activities for the maintenance of human health, including nutrient metabolism, epithelial cell proliferation, development and maintenance of the immune system, and protection against pathogens. Meanwhile, disruption of the normal microbiota (dysbiosis) has been linked to the various diseases such as obesity, cancer, a variety of inflammatory diseases of the airways, skin, mouth, and intestinal tract. To direct future

therapy for microbiome-related diseases, it is necessary to first characterize the “normal” microbiota in the healthy population and identify factors that are associated with compositional changes in human microbiome. For these purposes, in this thesis, a large-scale investigation of the gut microbiota and the airway microbiota was conducted in Korean twins and their families.

First, we identified that the gut microbiota of healthy Koreans were clustered into two enterotypes using metagenomic sequencing data which was generated from 36 fecal samples of healthy Korean monozygotic (MZ) twins. The two enterotypes were enriched by either *Bacteroides* or *Prevotella*. We observed that the enterotype of an individual was stable over time, and that most co-twins shared their enterotypes. We also found that the subject’s enterotypes were significantly associated with the serum uric acid level and the long-term intakes of nutrients such as dietary fiber, vitamins, and minerals. These results suggest that both host genetics and dietary habits could be important contributors to the enterotype of an individual.

Second, we characterized the gut microbiota of 655 MZ (N=306) and dizygotic (DZ) (N=74) twins, and family members of twin pairs (N=275), of which approximately 18% (121 individuals) had metabolic syndrome (MetS). We found that *Methanobrevibacter*, *Lactobacillus*, and *Sutterella* were significantly enriched in MetS individuals, while *Akkermansia*, *Odoribacter*, and *Bifidobacterium* were significantly enriched in healthy individuals. Among the gut microbes associated with MetS status, Actinobacteria, to which the *Bifidobacterium* belong, had the highest heritability (45.7%). Even after adjustment for MetS status, reduced abundances of Actinobacteria and *Bifidobacterium* were significantly linked to the minor allele at the apolipoprotein A-V gene (*APOA5*) SNP rs651821, which is associated with triglyceride level and MetS. Therefore, our results demonstrate that altered gut microbiota mediated by a specific host genotype can contribute to the development of MetS.

Third, we assessed the influences of host genetics and lifestyles such as smoking, alcohol consumption, and physical activity on the airway microbiota composition. A

total of 257 sputum samples from 74 MZ twin pairs (n=148), 14 DZ twin pairs (n=28), and their parents and siblings (n=81) were analyzed for airway microbiota composition using next-generation sequencing of partial 16S rRNA gene sequences. We found that several taxa, including *Providencia* and *Bacteroides*, were significantly influenced by host genetic factors. Smoking had the strongest effect on the overall microbial community structure among the various host lifestyle factors. The pack-year value was positively associated with abundance of the *Veillonella* genus, which is known to be related to airway inflammation, but negatively associated with that of *Haemophilus*. Co-occurrence network analysis showed that the taxa were clustered according to the direction of correlations or trends for smoking, and that the taxa influenced by host genetics were closely correlated with each other. These results demonstrate that the composition of the airway microbiota is shaped by complex interactions among host genetics and lifestyle factors, such as smoking.

In conclusion, this study shows that the composition of the healthy human microbiome is influenced not only by extrinsic environmental factors including diet and lifestyle but also by host genetics, and that the changes in the abundances of the specific microbial taxa caused by these factors are associated with metabolic syndrome and possibly airway inflammation. These data suggest that personalized approaches will be required in the prevention and treatment of microbiome-related diseases.

Keywords: Gut microbiome, airway microbiome, enterotype, metabolic syndrome, host genetics

***Student number*:** 2011-30706

CONTENTS

LIST OF TABLES

CHAPTER II.

CHAPTER III.

CHAPTER IV.

LIST OF FIGURES

CHAPTER I.

CHAPTER II.

CHAPTER III.

CHAPTER IV.

LIST OF ABBREVIATIONS

μL: microliter

μM: micromolar

APOA5: apolipoprotein A-V gene

BC: Bray-Curtis

CF: cystic fibrosis

SD: standard deviation

CH: Calinski–Harabasz pseudo F-statistic

COPD: chronic obstructive pulmonary disease

DNA: Deoxyribonucleic acid

DZ: dizygotic

EU: Euclidean

FBS: Fasting blood sugar

HDL: high-density lipoprotein cholesterol

HMP: Human Microbiome Project

JS: Jensen–Shannon

KEGG: Kyoto encyclopedia of genes and genomes

LDA: linear discriminant analysis

MetS: metabolic syndrome

min: minute

mL: milliliter

MZ: monozygotic

nM: nanomolar

NMDS: nonparametric multi-dimensional scaling

OTU: operational taxonomic unit

PAM: partitioning around medoids

PCR: polymerase chain reaction

QIIME: quantitative insights into microbial ecology

RDP: ribosomal database project

RNA: ribonucleic acid

S.E.M: standard error of the mean

SI: Rousseeuw's Silhouette internal cluster quality index

SNP: single nucleotide polymorphisms

CHAPTER I.

BACKGROUNDS

Human microbiome

Humans consist of not only our own somatic cells but also a tremendous diversity of commensal microbes including bacteria, archaea, fungi, and viruses. These microbes exist at all body surfaces that are exposed to the external environment, including the skin, oral cavity, airway, vagina, and gut, forming a community known as the human microbiome (Turnbaugh *et al.*, 2007). Over the past few years, with advances in sequencing methods and analytical techniques, the human microbiome has been intensively studied. As a representative example, the Human Microbiome Project (HMP) Consortium, the largest human microbiome study to date, has reported the structure and function of the 'normal' microbiome in 300 healthy US adults at 18 body sites using 16S rRNA gene sequencing (Human Microbiome Project Consortium, 2012). Through such efforts, our understanding of human microbiome biology have been greatly improved.

Each body site provides a highly specialized ecological niche for its own microbial community (Figure 1.1). For example, the gut microbiota is composed of members of Firmicutes and Bacteriodetes; the vaginal microbiota mainly comprises Firmicutes; the skin microbiota consists of Actinobacteria, Firmicutes, and Proteobacteria (Cho and Blaser, 2012). The existence of these body site-specific microbial communities imply that each habitat of the human body coevolved with specific microbes which harbor different functional repertoires (Pflughoeft and Versalovic, 2012). At each body site, members of the normal microbiome play critical roles in contributing to the maintenance of human health by participating in many metabolic functions, promoting differentiation of host tissues, modulating the immune system, or protecting the host against pathogens. In line with this, alterations in the composition of the human microbiome have been reported to be significantly associated with various diseases including obesity, autism, and a variety of inflammatory diseases of the airways, skin,

mouth, and intestinal tract (Cho and Blaser, 2012). Therefore, it is important to understand which factors contribute to compositional changes in human microbiome in order to prevent and control microbiome-related diseases.

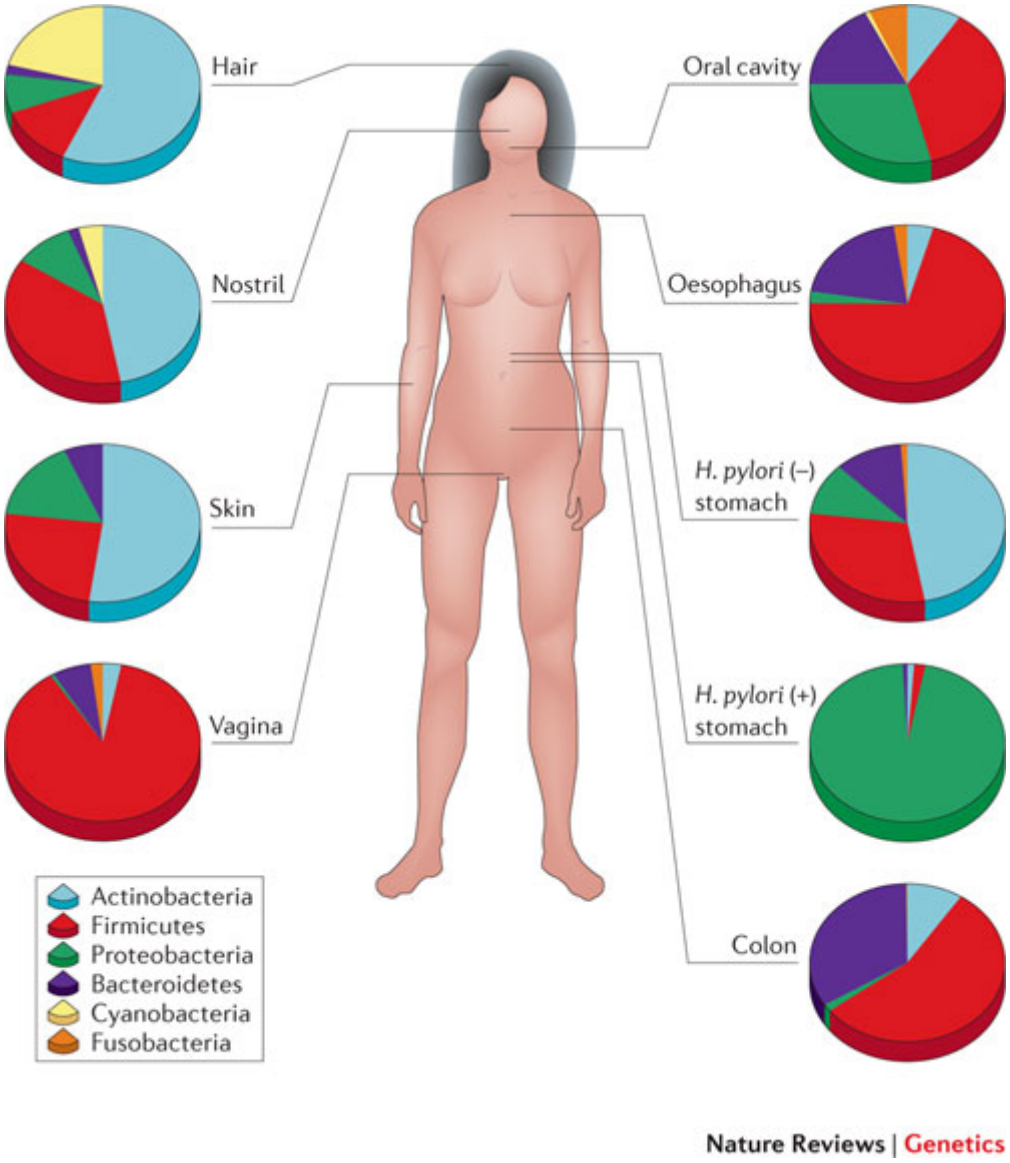


Figure 1.1 Compositional differences in the microbiome by anatomical site (Cho and Blaser, 2012).

Gut microbiome

In the human gut, a total of approximately 10^{14} microbial cell are present, which are tenfold as many cell as the rest of our bodies. Their collective genomes contain orders of magnitude more genes than our own genome (Gill *et al.*, 2006). The gut microbiota participates in a diverse range of biochemical and metabolic activities that are essential to the host. For example, these microbes break down indigestible dietary polysaccharides and produce micronutrients and vitamins; they regulate the function of the host's intestinal epithelium and immune system; they provide protection against invasion by potentially harmful pathogens (Sommer and Bäckhed, 2013).

The human intestine is a stable, protected, and nutrient-rich environment for the bacterial growth, and thus the composition of the adult gut microbiota is relatively stable over time. However, this habitat also can be changed due to antibiotic exposure, enteric infection, or a long-term change in diet, all of which can lead to alterations in the gut microbial composition. For example, in a healthy volunteer study, it has been shown that repeated antibiotic administration can result in the development of a new stable state in the gut microbiota (Dethlefsen and Relman, 2011).

Recent studies have revealed that there are significant variations in the composition of the gut microbiota of healthy individuals (Human Microbiome Project Consortium, 2012, Yatsunenko *et al.*, 2012). Several factors has been known to affect the gut microbial composition. For instance, aging is closely associated with changes in the gut microbiota. The relative abundance of *Bifidobacteria* was shown to be significantly decreased with increasing age in three different populations. Geography is another factor in shaping the gut microbiota. Large inter-personal variations was observed in the gut microbial communities of humans living in geographically and culturally distinct setting, indicating that the differences in host ethnicity, geographic location of residence, lifestyle, and diet may lead to the differences in the gut microbiota (Yatsunenko *et al.*, 2012). More recently, it has been reported that host genetics also influence the

composition of the microbiota (Goodrich *et al.*, 2014).

Taken together, the intra- and inter-personal variations driven by those various factors may be related to the development of the gut microbiome-related diseases such as obesity, diabetes, and inflammatory bowel diseases. Therefore, identifying specific disease-associated taxa and the factors that affect the abundances of such taxa will help optimize strategies to control disease.

Airway microbiome

The airways of healthy humans have been considered to be sterile when examined by culture-dependent techniques. However, recent studies using culture-independent techniques has revealed that even healthy airways harbor a unique steady-state microbiota (Charlson *et al.*, 2011, Erb-Downward *et al.*, 2011). Although the HMP did not include sampling of the lung, a series of studies to explore the airway microbiome is actively being conducted in similar ways to studies of other body sites. Accumulated data now indicate that the airway microbiota of healthy individuals is different from those of individuals with chronic lung diseases such as asthma, chronic obstructive pulmonary disease (COPD), and cystic fibrosis (Marsland and Gollwitzer, 2014), although whether changes in the microbiota are the cause or consequence of diseases remains to be elucidated.

The respiratory tract is associated not only with antimicrobial peptide-producing epithelial cells but also with various immune cells including antibody-producing B cells, effector or memory T cells, dendritic cells. Accordingly, the airway microbiota can play critical roles in modulation of the immune response in the respiratory tract, tightly interacting with host factors. For example, studies in childhood asthma have shown that exposure to a diverse microbial environment early in life is protective against asthma (Ege *et al.*, 2011, Fujimura *et al.*, 2010), emphasizing the importance of the airway microbiota in the development of the immune system.

Given the clear differences in the airway microbiota between healthy and diseased individuals, it is important to characterize the “normal” airway microbiota in the healthy population for valid comparison. Much remains to be studied regarding the factors that underlie alterations in the composition of the airway microbiota and their relationships with specific microbes for the development of patient-tailored strategies for lung disease.

Techniques for studying the human microbiome

Advances in low-cost high-throughput sequencing techniques enabled large-scale surveys of the human microbiome such as the HMP (Turnbaugh *et al.*, 2007) and the Metagenomics of the Human Intestinal Tract (MetaHIT) consortium (Qin *et al.*, 2010). These rapid advances also led to the development of diverse analytic techniques and tools.

16S ribosomal RNA gene (16S) sequencing is the most commonly used technique for studying the human microbiome. This allows researcher to rapidly and cost-effectively identify the bacteria and archaea that are present in the communities of interest. However, this method has several limitations. 16S sequencing cannot identify viral and fungal communities. Also, 16S sequencing cannot directly identify the metabolic and functional capabilities of the microbial community. To predict community-wide functional profiles using 16S sequence data and a database of reference genomes, a recent bioinformatics pipeline, PICRUSt, can be used (Langille *et al.*, 2013).

Metagenome sequencing is a technique to sequence the DNA of an entire microbial community including archaea, bacteria, fungi, and viruses. The resulting metagenomic data can yield species-level taxonomic profiles using bioinformatic tools such as MetaPhlAn (Segata *et al.*, 2012), and detailed metabolic and functional profiles using pipelines such as HUMAnN (Abubucker *et al.*, 2012). However, deep metagenomic sequencing for many samples is still expensive, and computational challenges in processing sequence data remain formidable.

Taxonomic and functional profiles produced from 16S or metagenomic sequence data are used for subsequent visualization and analysis. The diversity of microbial communities is traditionally described in terms of alpha (or within-sample) diversity and beta (or between-sample) diversities. Alpha diversity is measured as the number and distribution of taxa expected within a given community. On the other hand, beta diversity is measured as the degree to which samples differ from one another, allowing

further analysis by dimensionality reductions, such as principal coordinates analysis (PCoA), or by sample clustering. Clustering of human gut microbiome samples based on their taxonomic similarity led to the concept of enterotype, which can be used as a way of stratifying samples to reduce the considerable variation in the composition of the human microbiome. In addition, specific taxa or functional pathways which are significantly associated with metadata of interest can be identified by tools such as LEfSe (Segata *et al.*, 2011) or MaAsLin (Morgan *et al.*, 2012).

These sequencing techniques and analytic tools have allowed us to gain a more accurate understanding of the human microbiota and its functions in human health and disease (Figure 1.2).

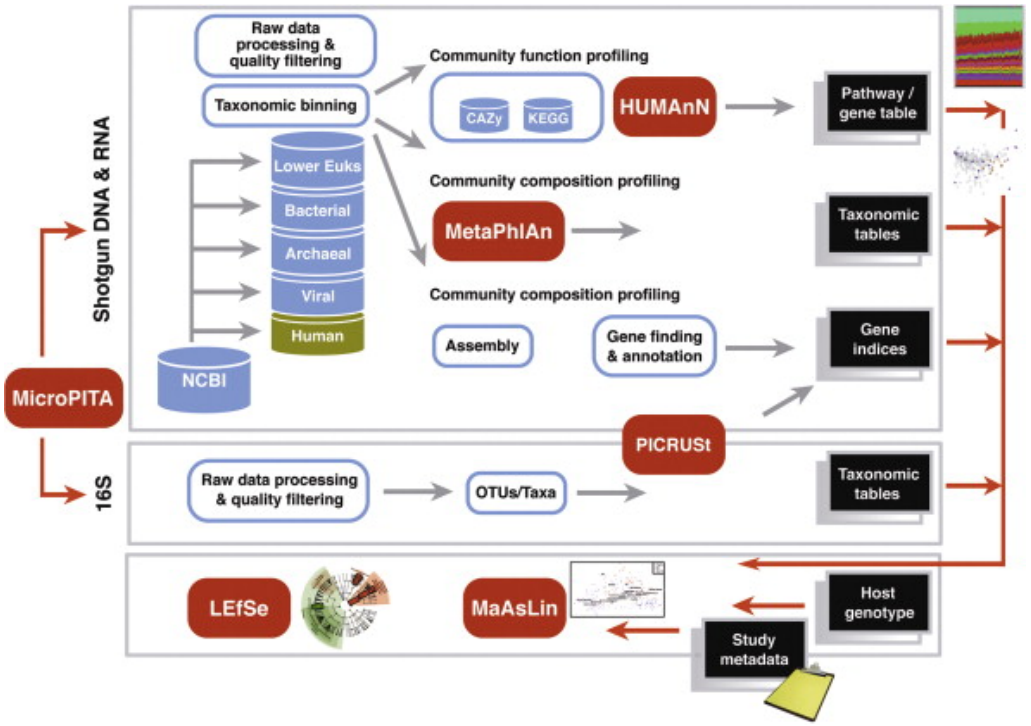


Figure 1.2 A workflow for studying the human microbiome (Morgan and Huttenhower, 2014)

Objectives and Hypotheses

Objectives:

The objectives of this study were 1) to identify the enterotypes of healthy Koreans and determine their characteristics, 2) to investigate links between the gut microbiota, host genetics, and metabolic syndrome, and 3) to assess the influence of host lifestyle and genetic factors on the healthy airway microbiota in Korean twins and their families.

Hypothesis No. 1:

The gut microbiota of healthy Koreans can be categorized into enterotypes, and specific enterotypes are associated with host-related factors.

Hypothesis No. 2:

There are differences in the composition of the gut microbiota between healthy individuals and individuals with metabolic syndrome, and host genetic factors can partially contribute to the differences.

Hypothesis No. 3:

The composition of the healthy airway microbiota is shaped by both lifestyle factors and host genetic factors.

CHAPTER II.

STABILITY OF GUT ENTEROTYPES IN KOREAN MONOZYGOTIC TWINS AND THEIR ASSOCIATION WITH BIOMARKERS AND DIET

Introduction

Trillions of microbes reside in the human gut, carrying out functions that are critical to host health, including dietary energy extraction, development and maintenance of the immune system, and protection against pathogens (Candela *et al.*, 2008, Cebra, 1999, Turnbaugh *et al.*, 2006). The composition of the gut microbiota is known to vary greatly among healthy individuals (Human Microbiome Project Consortium, 2012, Turnbaugh *et al.*, 2009), and to be influenced by various factors such as age, nutrition, and geography (De Filippo *et al.*, 2010, Yatsunenko *et al.*, 2012). In addition, compositional and functional changes in the gut microbiota have been observed in patients with several chronic disorders including obesity, type 2 diabetes, and inflammatory bowel disease (Morgan *et al.*, 2012, Qin *et al.*, 2012, Turnbaugh *et al.*, 2006).

Arumugam *et al.* suggested that human gut microbial communities can be clustered into three enterotypes, each of which is represented by different dominant genera: *Bacteroides*, *Prevotella*, or *Ruminococcus*. These enterotypes were not significantly correlated with host properties such as nationality, age, gender, or body mass index (BMI) (Arumugam *et al.*, 2011). Subsequent studies have demonstrated associations of enterotypes with long-term dietary patterns (Wu *et al.*, 2011) and several diseases such as symptomatic atherosclerosis (Karlsson *et al.*, 2012) and nonalcoholic steatohepatitis (Zhu *et al.*, 2013). Recently, it has been somewhat controversial about the number and discreteness of enterotypes (Claesson *et al.*, 2012, Koren *et al.*, 2013, Ou *et al.*, 2013, Wu *et al.*, 2011). A recent study using Human Microbiome Project (HMP) data showed that the gut microbiota is represented as a bimodal distribution and smooth gradient rather than a discrete distribution (Koren *et al.*, 2013). More recently, Arumugam *et al.* stated in the addendum of their original enterotype paper that enterotypes should be considered as a way to simplify the complexity of the gut microbiota rather than as distinct clusters (Arumugam *et al.*, 2014). So far, however, it is unknown whether the

enterotypes that were studied mainly in populations from Europe and the United States (Arumugam *et al.*, 2011, Huse *et al.*, 2012, Wu *et al.*, 2011) are present in an Asian population with different dietary and genetic backgrounds. In addition, whether enterotypes are influenced by host genotype, or how stable enterotypes are over time, remains uncertain.

Here, we conducted taxonomic and functional profiling of gut microbiomes from healthy Korean monozygotic (MZ) twins to characterize their enterotypes and functional clusters, and to determine the presence or absence of genetic effects on enterotypes as well as the longitudinal stability of enterotypes in a Korean population. In addition, we further investigated associations of enterotypes with long-term dietary habits and clinical biomarkers.

Materials and Methods

Sample collection

Thirty-six fecal samples were obtained from 20 MZ twins (10 MZ twin pairs) enrolled in the Healthy Twin Study in Korea (Sung *et al.*, 2006). Among them, 16 MZ twins (8 MZ twin pairs) provided fecal samples at two time points with an average interval of 2 years, and 4 MZ twins (2 MZ twin pairs) provided samples at only one time point. Therefore, 36 fecal samples were collected and subsequently analyzed. The fecal samples were produced at home and immediately placed in freezer. Frozen fecal samples were brought to clinics and stored at –80°C prior to subsequent analysis. Antibiotic exposure was considered exclusion criteria in this study.

Metadata collection

Anthropometrical measurements and biochemical tests were conducted for each subject at every visit. Weight and height were measured using the InBody 3.0 (Biospace Inc., Seoul, Korea) body composition analyzer, and BMI was calculated as the weight (kg) divided by the square of the height (m^2). Blood pressure was measured with a standard mercury sphygmomanometer. Blood samples were drawn from an antecubital vein after an overnight fast of at least 8 h and sent to one central laboratory for blood tests within 24 h. Measurements of glucose, total cholesterol, triglyceride, high-density lipoprotein, and low-density lipoprotein were assessed by enzymatic methods. Insulin levels were measured by a radioimmunoassay method. The serum uric acid level was assayed by a uricase enzymatic colorimetric method, and high-sensitive C-reactive protein was measured by latex agglutination.

Participants also filled out questionnaires covering lifestyle, medication, and disease history. Long-term dietary information was obtained using a 106-item food frequency questionnaire for Koreans (Ahn *et al.*, 2007). For each food item, the intake in grams

per day was estimated based on the consumption frequency (nine categories: never or seldom, once a month, two to three times a month, once or twice a week, three to four times a week, five to six times a week, once a day, twice a day, or three times or more every day) and the portion size (three categories: small, medium, or large). Average daily intakes of energy and 23 nutrients including protein, fat, carbohydrates, fiber, vitamins, and minerals were calculated using a computer program (Korean Genomic Epidemiological Cohort Study Information System, version 1.0), but this information was missing for four of the 36 samples. For the analyses, we used the average daily intake for each of the 30 food groups that were created by combining similar food items into a single group. We also used energy-adjusted nutrient intakes that were calculated as the standardized residuals from the regression of a specific nutrient on energy.

All experiments were performed in accordance with relevant guidelines and regulations approved by the institutional review board of Samsung Medical Center, Busan Paik Hospital, and Seoul National University (IRB No. 144-2011-07-11). Written informed consent was obtained from all participants.

Metagenomic shotgun sequencing and analysis

DNA extraction from each fecal sample was conducted using the MoBio Power Soil DNA Isolation Kit (MoBio, Solana Beach, CA, USA). All samples were sequenced by 101-bp paired-end sequencing on a HiSeq platform (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. The raw reads were filtered to remove low-quality reads and human sequence contamination. To estimate the taxonomic composition of each sample, the filtered reads were mapped to a set of clade-specific marker genes using MetaPhlAn (Segata *et al.*, 2012). For metabolic reconstruction, high-quality reads were mapped to the protein-coding sequences from the KEGG Orthology using USEARCH, and best hits were further passed through HUMAnN (Abubucker *et al.*, 2012).

Enterotyping and functional clustering

For enterotyping and functional clustering of the gut microbiota, we calculated Jensen–Shannon (JS) distance for the genus-level relative abundance profiles and for the KEGG pathway relative abundance profiles. Recently, Koren *et al*. recommended using at least two distance metrics because of the sensitivity of enterotyping to the distance metrics (Koren *et al.*, 2013). Therefore, we also applied the two other distance metrics (Bray-Curtis [BC] and Euclidean [EU] distances). To cluster the samples based on these distance metrics, we used the partitioning around medoids (PAM) method in the R package 'cluster'. To determine the optimal number of clusters and evaluate the cluster quality, we calculated Calinski–Harabasz pseudo F-statistic (CH) and Rousseeuw's Silhouette internal cluster quality index (SI) values for each of three distance metrics as a relative and an absolute measure, respectively, recommended by Koren *et al.* (Koren *et al.*, 2013) using the R package 'clusterSim'. The samples were plotted by principal coordinate analysis using the R package 'ade4'.

Statistical analysis

To determine taxonomic and metabolic features that were differentially abundant either between enterotypes or between functional clusters, LEfSe (linear discriminant analysis [LDA] effect size) was applied under the condition $\alpha = 0.01$, with an LDA score of at least 2 (Segata *et al.*, 2011). The concordance rate for the enterotype in MZ twins was calculated as the percentage of paired fecal samples from MZ twin pairs that were assigned to the same enterotype at each time point. Similarly, the persistency rate of the enterotype over time was calculated as the percentage of individuals who were assigned to the same enterotype at two different time points. Two-sample permutation tests were conducted on JS distances of the gut microbiota based on the genus-level relative abundances for the four types of comparisons for all samples obtained from each individual: an individual at different time points, twin pairs at the same time point, twin pairs at different time points, and unrelated individuals. Differences in

anthropometrical/biochemical biomarker measures and long-term diets (both dietary patterns and nutritional factors) between enterotypes were analyzed using the nonparametric Wilcoxon rank-sum test. To identify host properties that demonstrated significantly different abundances between co-twins discordant for enterotype, we applied the Wilcoxon signed-rank test. Heat maps of the energy-adjusted nutrient intakes across samples were generated using the Multi-Experiment Viewer software (version 4.8.1). Hierarchical clustering of samples was performed with the energy-adjusted nutrient intake data using EU distance and an average linkage method. Fisher's exact test was used to determine associations between enterotypes and groups clustered based on long-term patterns of nutrient intakes. P-values were adjusted for multiple testing with the Benjamini–Hochberg method.

Results

Clustering Korean gut microbiome into enterotypes

We investigated whether the gut microbiota of healthy Korean MZ twins could be grouped into enterotypes, following the procedure of the previous study (Arumugam *et al.*, 2011). The characteristics of participants in the study are summarized in Table 2.1. The highest CH value was obtained for two clusters for all of the tested distance metrics. For two clusters, SI provided weak support ($0.25 < SI \leq 0.5$) using EU distance, and no support ($SI \leq 0.25$) using JS and BC distances by the criteria proposed by Kaufman and Rousseeuw (Kaufman and Rousseeuw, 2009) (Figure 2.1). Interestingly, we observed that the sample assignments to clusters obtained by using JS and EU distances were perfectly identical (Table 2.2). We thus inferred that these two distance metrics formed reliable clusters from our metagenome data, and chose to use the two clusters generated using JS distance for further analyses, as in Arumugam *et al*. (Arumugam *et al.*, 2011).

Table 2.1 Characteristics of participants in this study.

	n (%) or mean±SD		
	Time point 1	**Time point 2**	**Total**
	n=20	**n=16**	**n=36**
Age (years)	38.9±6.5	39±6.3	38.9±6.3
Male:Female	8:12	6:10	14:22
Body mass index (kg/m^2)	23.6±3.7	23.1±3.4	23.4±3.5
Fasting blood sugar (mg/dL)	93.5±12.6	96.4±15	94.8±13.6
Uric acid (mg/dL)	4.9±1.8	4.7±1.9	4.8±1.8
Triglyceride (mg/dL)	124.9±136.2	109.3±39.6	117.9±103.9
Total cholesterol (mg/dL)	184.7±29.5	192.3±45.1	188.1±36.8
Systolic blood pressure (mmHg)	113.7±15.5	115.1±16	114.3±15.5
Diastolic blood pressure (mmHg)	71.3±10.1	75.4±11.6	73.1±10.9
HDL-cholesterol (mg/dL)	47.1±10.7	41.5±8.9	44.6±10.2
LDL-cholesterol (mg/dL)	113.8±26.7	118.8±35.2	116±30.4
Waist circumference (cm)	79.2±9.9	78.3±11.3	78.8±10.4

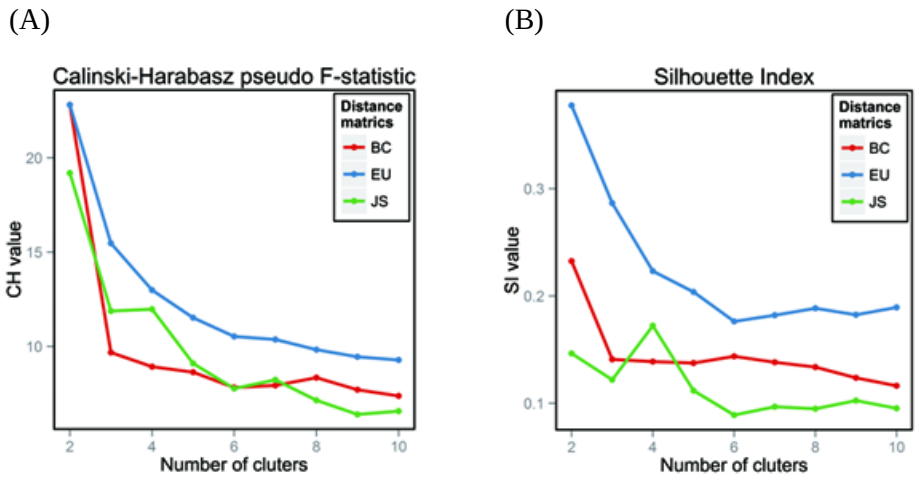


Figure 2.1 Estimation of the optimal number of enterotypes. (A) The Calinski–Harabasz pseudo F-statistic and (B) Rousseeuw's Silhouette internal cluster quality index for three distance metrics (Jensen–Shannon [JS], Bray–Curtis [BC], Euclidean [EU] distances) of the genus-level relative abundance profiles.

Table 2.2 Sample assignment to the enterotypes for three distance metrics (Jensen–Shannon [JS], Bray–Curtis [BC], Euclidean [EU] distances) and to the functional clusters based on JS distance.

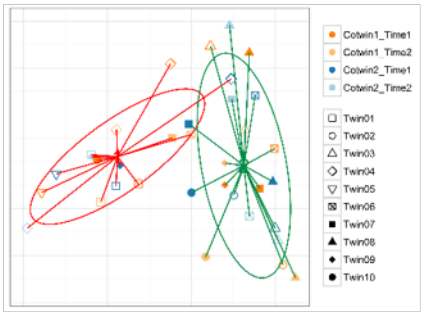
SampleID	Twin	Enterotype			Functional cluster
		JS	EU	BC	JS
S183_1	01	1	1	1	2
S183_2	01	1	1	1	1
S184_1	01	1	1	1	1
S184_2	01	1	1	1	1
S197_2	02	2	2	2	2
S198_2	02	2	2	2	2
S205_1	03	2	2	2	2
S205_2	03	2	2	2	2
S206_1	03	2	2	2	2
S206_2	03	2	2	2	2
S223_1	04	1	1	1	2
S223_2	04	1	1	1	2
S224_1	04	1	1	2	2
S224_2	04	1	1	1	1
S241_2	05	1	1	1	1
S242_2	05	1	1	1	1
S249_1	06	2	2	2	2
S249_2	06	1	1	2	2
S250_1	06	2	2	2	2
S250_2	06	2	2	2	2
S251_1	07	2	2	2	2
S251_2	07	1	1	2	2
S252_1	07	2	2	2	2
S252_2	07	2	2	2	2
S267_1	08	2	2	2	2
S267_2	08	2	2	2	2
S268_1	08	2	2	2	2
S268_2	08	2	2	2	2
S279_1	09	2	2	2	2
S279_2	09	2	2	2	2
S280_1	09	1	1	1	2
S280_2	09	1	1	1	1
S281_1	10	1	1	1	1
S281_2	10	2	2	2	2
S282_1	10	2	2	2	2
S282_2	10	2	2	2	2

Compositional and functional characteristics of enterotypes

Fifteen (41.7%) of the 36 samples were assigned to enterotype 1, and the rest (58.3%) were assigned to enterotype 2 (Figure 2.2A, B and Table 2.2). To identify the genera with different abundance values between the two enterotypes, LEfSe analysis was performed. In agreement with previous studies (Claesson *et al.*, 2012, Wu *et al.*, 2011), enterotype 1 and 2 were enriched in *Bacteroides* and *Prevotella*, respectively (Figure 2.2C). We also observed that *Catenibacterium* was overrepresented in enterotype 1, whereas 23 genera, including *Lactobacillus*, *Dorea*, and *Coprococcus*, were more abundant in enterotype 2 (Figure 2.2D).

To investigate the functional potential of each enterotype, we applied LEfSe to the relative abundance profiles of KEGG modules. The functional modules involved in the biosynthesis of vitamins, such as riboflavin (vitamin B_2), pantothenate (vitamin B_5), biotin (vitamin B_7), and tetrahydrofolate (vitamin B_9 derivative), were highly abundant in enterotype 1. Additionally, modules for pentose phosphate pathways and nucleotide (guanine and adenine) biosynthesis were also enriched in enterotype 1 (Figure 2.3). On the other hand, enterotype 2 was associated with an increased abundance of transport systems for glutathione, multiple sugars, and branched chain amino acids.

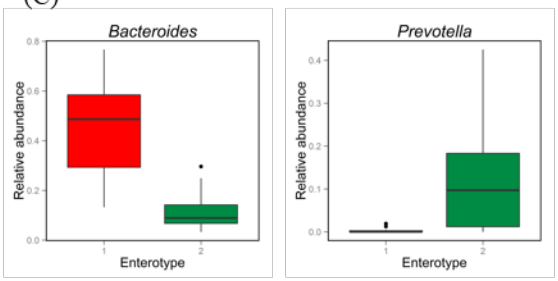
(A)


Cotwin1_Time1
Cotwin1_Time2
Cotwin2_Time1
Cotwin2_Time2
Twin01
Twin02
Twin03
Twin04
Twin05
Twin06
Twin07
Twin08
Twin09
Twin10

(B)


Twin # 01 02 03 04 05 06 07 08 09 10
Time point 1
Cotwin 1
S183 S197 S205 S223 S241 S249 S251 S267 S279 S281
Cotwin 2
S184 S198 S206 S224 S242 S250 S252 S268 S280 S282
Time point 2
Cotwin 1
Cotwin 2

(C)


Bacteroides
Relative abundance
Enterotype
Prevotella
Relative abundance
Enterotype

(D)

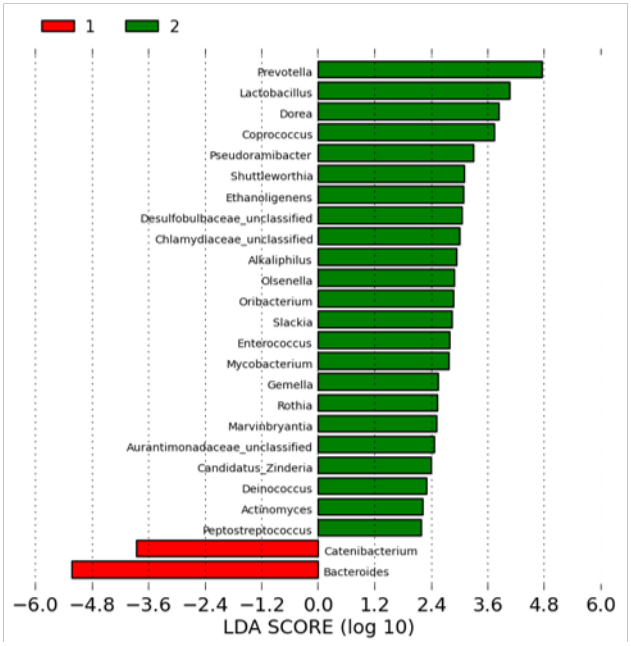

1
2
Prevotella
Lactobacillus
Dorea
Coprococcus
Pseudoramibacter
Shuttleworthia
Ethanoligenens
Desulfobulbaceae_unclassified
Chlamydiaceae_unclassified
Alkaliphilus
Olsenella
Oribacterium
Slackia
Enterococcus
Mycobacterium
Gemella
Rothia
Marvinbryantia
Aurantimonadaceae_unclassified
Candidatus_Zinderia
Deinococcus
Actinomyces
Peptostreptococcus
Catenibacterium
Bacteroides
−6.0 −4.8 −3.6 −2.4 −1.2 0.0 1.2 2.4 3.6 4.8 6.0
LDA SCORE (log 10)

Figure 2.2 Identification of enterotypes in a Korean population. (A) The first two principal coordinates of the Jensen–Shannon distances of the genus-level relative abundance profiles. Samples are colored by enterotype as identified by the partitioning around medoids (PAM) clustering algorithm. Red is enterotype 1 and green is enterotype 2. (B) Sample assignment to the enterotypes. Samples assigned to enterotype 1 are represented by red, samples assigned to enterotype 2 are represented by green. (C) Relative abundances of *Bacteroides* and *Prevotella* in each enterotype. Boxes represent the interquartile range (IQR) between the first and third quartiles, with a line at the median. (D) Histogram of the linear discriminant analysis (LDA) score for differentially abundant genera between the enterotypes. Negative (red bars) and positive (green bars) LDA scores represent genera overrepresented in enterotype 1 and enterotype 2, respectively. Features with LDA scores >2 are presented.

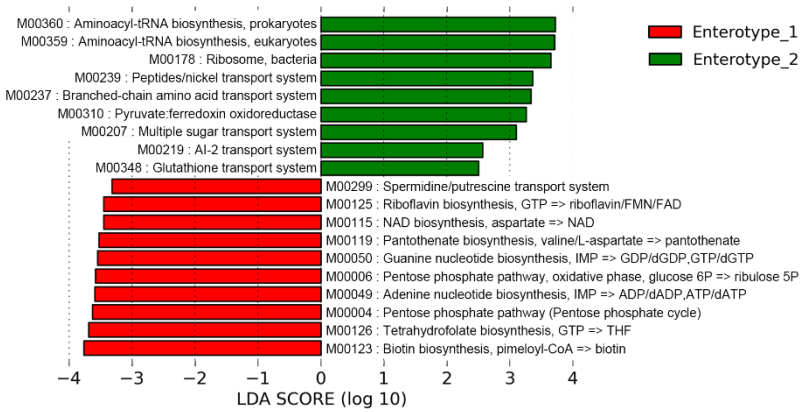


Figure 2.3 Functional differences between enterotypes. Histogram of the linear discriminant analysis (LDA) scores for differentially abundant KEGG modules between the enterotypes. Negative (red bars) and positive (green bars) LDA scores represent KEGG modules overrepresented in enterotype 1 and enterotype 2, respectively. Features with LDA scores >2 are presented.

Genetic and temporal influences on enterotypes

Our subjects consisted of MZ twin pairs, some of whom had longitudinal samples collected after an average interval of 2 years. To determine the genetic and temporal influences on enterotypes, we subsequently calculated the concordance rate for the enterotype in MZ twins and the persistency rate of the enterotype. In terms of the concordance rate between the MZ co-twins, 13 (72.2%) of the 18 paired fecal samples from 10 MZ twin pairs (2 MZ pairs at one time point and 8 MZ twin pairs at two time points) belonged to the same enterotype. The kappa measure of agreement for the paired fecal samples was $\kappa = 0.43$. Five and eight paired fecal samples from 9 MZ twin pairs belonged to enterotype 1 and enterotype 2, respectively. Among the 18 paired fecal samples, 5 paired fecal samples from 4 MZ twin pairs (3 MZ twin pairs at one time point and 1 twin pair at two time points) were discordant for enterotype, meaning that one sibling of each twin pair was placed in enterotype 1 and the other sibling was placed in enterotype 2 (Figure 2.2B and Table 2.3). In terms of the persistency of the enterotype of the same individual after an average of 2 years, 13 (81.3%) of the 16 individuals were assigned to the same enterotype, whereas only 3 (18.7%) of the 16 individuals showed enterotype changes over the 2-year period ($\kappa = 0.61$) (Figure 2.2B and Figure 2.4).

To confirm our observation, we additionally compared the JS distances between individuals over time, MZ twin pairs (at the same time point and at two different time points), and unrelated individuals (Figure 2.5). The results indicated that the JS distance between communities from co-twins at the same time point was shorter than the distance between communities from unrelated individuals, bordering on statistical significance ($P = 0.0765$). We also observed that the same individual had significantly more similar gut microbial communities over about a 2-year interval compared to those from unrelated individuals ($P = 0.0015$).

Table 2.3 Enterotype assignment of the paired fecal samples from the MZ twin pairs at each of the two time points

Enterotype	Concordant paired fecal samples		Discordant paired fecal samples		Total	
	N	%	N	%	N	%
1	5	27.8	-	-	5	27.8
2	8	44.4	-	-	8	44.4
1/2	-	-	5	27.8	5	27.8
Total	13	72.2	5	27.8	18	100.0

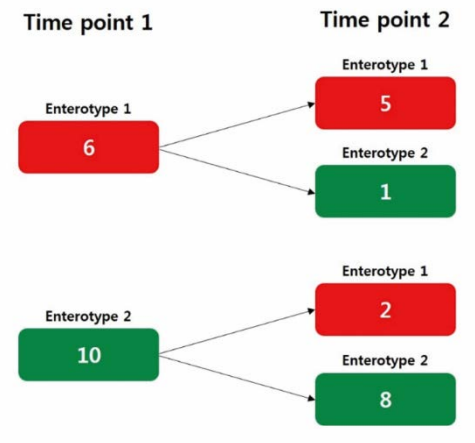


Figure 2.4 Persistency of enterotypes over time. The numbers of samples assigned to each enterotype at two different time points over the 2-year period are presented.

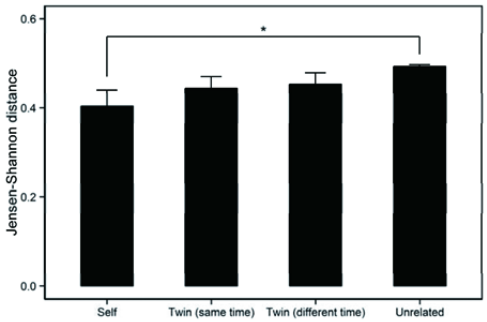


Figure 2.5 Jensen–Shannon distances of the genus-level relative abundance profiles between individuals over time (self), twin pairs (at the same time point and at two different time points), and unrelated individuals (two sample permutation test; *P < 0.05; mean ± s.e.m.).

Functional clusters based on KEGG pathway profiles

We further investigated the characteristics of functional clusters based on KEGG pathway abundances. The optimal number of clusters was two, when the JS distance metric was applied (Figure 2.6). All samples that were classified as enterotype 2 were grouped into functional cluster 2. However, among the 15 fecal samples classified as enterotype 1, only eight were assigned to functional cluster 1, which was comprised of only enterotype 1 samples. The other seven enterotype 1 samples were assigned to functional cluster 2 along with all of the enterotype 2 samples (Table 2.2 and Table 2.4). These seven samples had relatively low levels of *Bacteroides* and higher levels of genera that were overrepresented in enterotype 2 compared to the rest of the enterotype 1 samples, with an alpha value of 0.05 based on LEfSe (data not shown). This finding suggests that some samples belonging to enterotype 1 appeared to be functionally closer to enterotype 2 than to enterotype 1.

When we performed LEfSe analysis to identify differentially abundant KEGG modules between two functional clusters, we observed that all of the KEGG modules overrepresented in enterotype 1 were also enriched in functional cluster 1 (Figure 2.7). In functional cluster 2, several additional KEGG modules were enriched, such as threonine and lysine biosynthesis, the maltose/maltodextrin and putative sugar transport systems, and the cobalt transport system, which were not enriched in enterotype 2 based on genus-level abundances. Overall, our results show that the enterotypes based on genus-level abundances do not necessarily correspond with the functional clusters based on metabolic pathway profiles.

(A) (B)

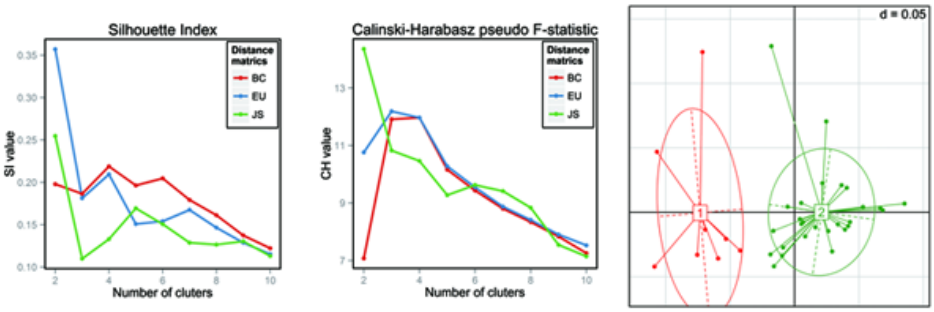


Figure 2.6 Identification of two functional clusters based on KEGG pathway profiles. (A) Estimation of the optimal number of functional clusters calculated from the Calinski–Harabasz pseudo F-statistic and Rousseeuw's Silhouette internal cluster quality index for three distance metrics (Jensen–Shannon [JS], Bray–Curtis [BC], Euclidean [EU] distances) of the KEGG pathway relative abundance profiles, and (B) the first two principal coordinates of the Jensen–Shannon distances of the KEGG pathway profiles. Samples are colored by functional cluster as identified by the partitioning around medoids (PAM) clustering algorithm. Red is functional cluster 1 and green is functional cluster 2.

Table 2.4 Distribution of samples based on enterotypes and functional clusters

Metabolic pathway	Community composition				Total	
	Enterotype 1		Enterotype 2			
	N	%	N	%	N	%
Functional cluster 1	8	22.2	0	0.0	8	22.2
Functional cluster 2	7	19.4	21	58.3	28	77.8
Total	15	41.7	21	58.3	36	100.0

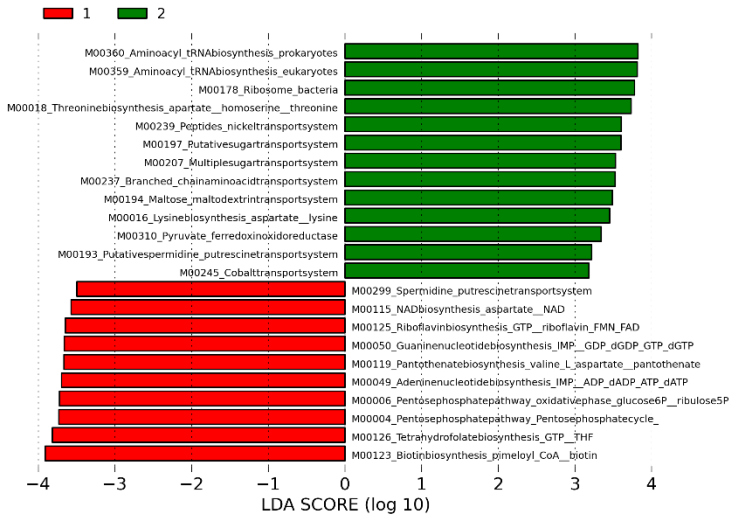


Figure 2.7 Functional differences between two functional clusters based on KEGG pathway profiles. Histogram of the linear discriminant analysis (LDA) scores for differentially abundant KEGG modules between the functional clusters. Negative (red bars) and positive (green bars) LDA scores represent KEGG modules overrepresented in functional cluster 1 and functional cluster 2, respectively. Features with LDA scores >2 are presented.

The associations of enterotypes with clinical biomarkers

We investigated the association of enterotypes with host properties including anthropometrical/biochemical measures. The enterotypes were not significantly correlated with biomarkers such as age, BMI, blood pressure, fasting blood sugar, total cholesterol, or triglyceride (Figure 2.8). The only biomarker found to be significantly different between the enterotypes was the serum uric acid level (Wilcoxon rank-sum test, adj. $P = 0.04$). Enterotype 2 had significantly higher levels of uric acid than enterotype 1 (Figure 2.8).

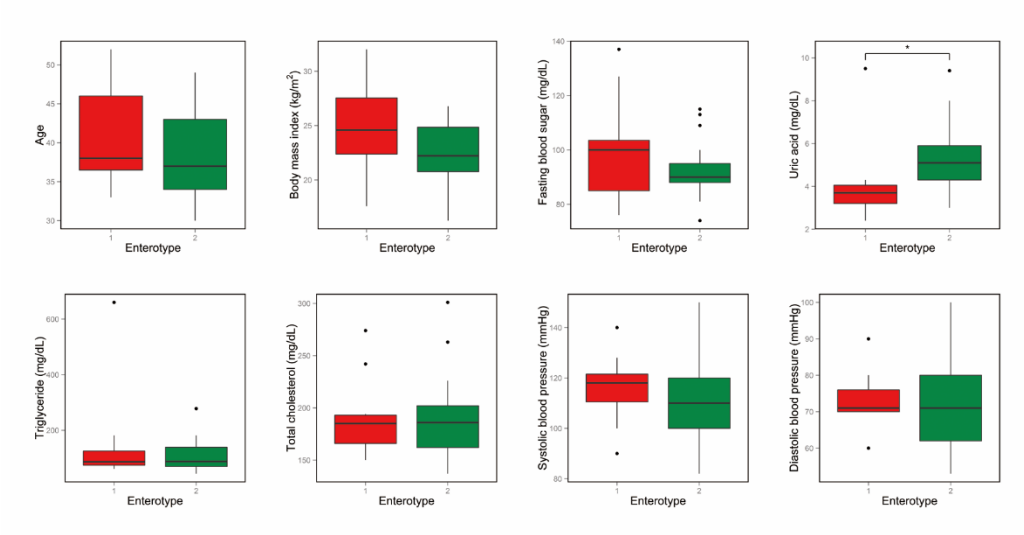


Figure 2.8 Association of enterotypes with host properties. The Wilcoxon rank-sum test was used to assess the association of enterotypes with age, body mass index, and clinical biomarkers (*adj. $P < 0.05$). Boxes represent the interquartile range (IQR) between the first and third quartiles, with a line at the median.

The associations of enterotypes with long-term diets

We next investigated associations between enterotype and long-term diets using the Wilcoxon rank-sum test. Enterotype 2 was associated with diets higher in fruit and egg food items compared to enterotype 1 (Figure 2.9A; adj. P = 0.03). For energy-adjusted nutrient intakes, enterotype 2 had significantly higher levels of dietary fiber, minerals (potassium and iron), and vitamins (vitamin A, vitamin C, vitamin E, folate, carotene, and retinol) than enterotype 1 (Figure 2.9B). No significant associations were noted between enterotypes and macronutrients such as protein, fat, and carbohydrate (Figure 2.9B).

We next examined the association of overall long-term dietary patterns with the enterotypes. Hierarchical clustering revealed that samples could be clustered into two groups based on 23 energy-adjusted nutrient values (Figure 2.10). However, no statistically significant association between two clustered groups and enterotypes was observed. On the other hand, when we performed hierarchical clustering using only 9 energy-adjusted nutrient values related to the enterotypes (Figure 2.9C), we could observe not only the clustering of the samples into two groups, but also the significant association between these groups and the enterotypes (Fisher's exact test, P = 0.03). These results suggest that the major determinants of the enterotypes are the long-term intakes of dietary fiber and micronutrients, rather than protein, fat, or carbohydrate.

Finally, for enterotype-discordant twin pairs, we performed the Wilcoxon signed-rank test to identify significantly different host characteristics including anthropometrical, biochemical, and dietary features. No significant difference was found at a false discovery rate of 5%. Similarly, no significant differences were observed in host characteristics between functional clusters based on KEGG pathway profiles.

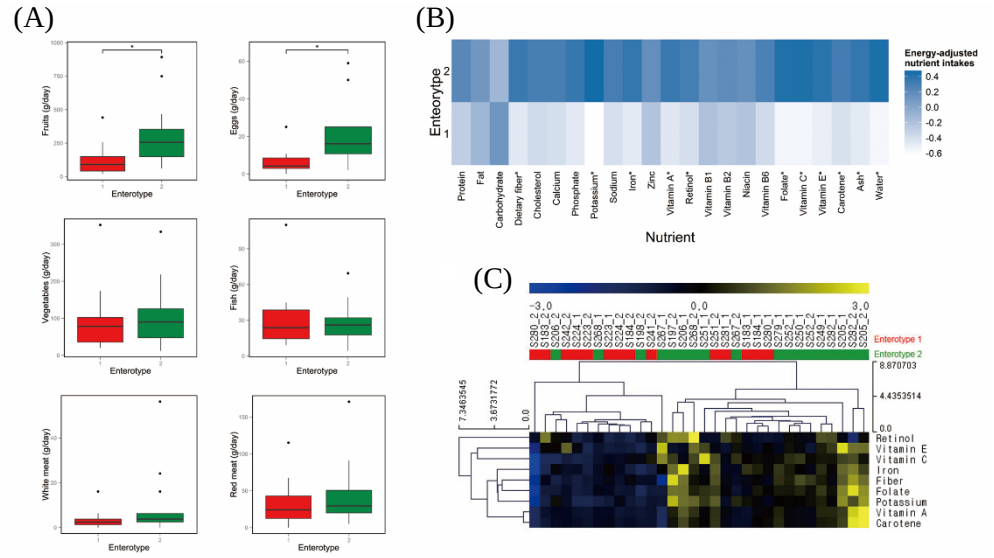


Figure 2.9 Association of enterotypes with long-term dietary intakes. (A)-(B), The Wilcoxon rank-sum test was used to assess the association of (A) intakes of food items and (B) energy-adjusted nutrient intakes with the enterotypes (*adj. $P < 0.05$). In (A), boxes represent the interquartile range (IQR) between the first and third quartiles, with a line at the median. In (B), colors represent the mean values of the standardized residuals obtained from the regression of a specific nutrient on energy. Darker blue corresponds to higher intake of the nutrient. (C) Heat map of hierarchical clustering for only the nutrient variables shown to be significantly associated with the enterotypes in this study. Clustering was performed using Euclidean distance and an average linkage method. Red indicates enterotype 1 and green indicates enterotype 2.

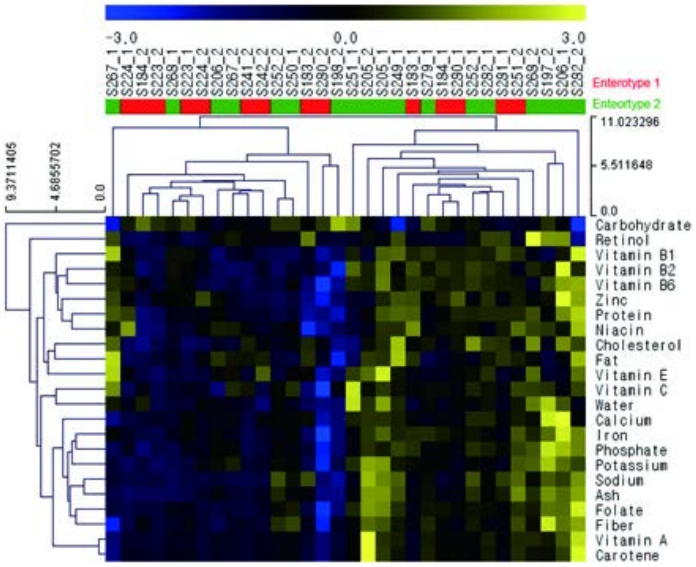


Figure 2.10 Heat map of hierarchical clustering for 23 energy-adjusted nutrient values. Clustering was performed using Euclidean distance and an average linkage method. Red indicates enterotype 1 and green indicates enterotype 2.

Discussion

The enterotypes of human gut microbiota have recently been studied in populations in Europe and the United States, and the number of enterotypes was somewhat controversial (Arumugam *et al.*, 2011, Claesson *et al.*, 2012, Ou *et al.*, 2013, Wu *et al.*, 2011). Our study clearly indicated that healthy Koreans, with completely different ethnic group and diet, are categorized into two different enterotypes based on their gut microbial composition: *Bacteroides* (enterotype 1) and *Prevotella* (enterotype 2). The third enterotype with high abundances of *Ruminococcus*, suggested in the previous study by Arumugam *et al.* (Arumugam *et al.*, 2011), was not observed in this study. A study using HMP data showed that the sample size could affect the number of enterotypes (Zhou *et al.*, 2014). In this regard, our results should be interpreted with caution due to the relatively small sample size. However, our results were consistent with a number of previous studies with larger sample sizes. In these studies, two enterotypes were also observed (Claesson *et al.*, 2012, Wu *et al.*, 2011), and a recent study by Koren *et al*. indicated the bimodal distribution of gut microbiota (Koren *et al.*, 2013). Therefore, our results confirm that most humans can be clustered into at least two enterotypes, regardless of the ethnicity or geographic region (Arumugam *et al.*, 2011).

Our analysis showed that 58.3% (n=21) of samples belonged to *Prevotella* enterotype, whereas 41.7% (n=15) of samples belonged to *Bacteroides* enterotype. In contrast, it has been reported that in Western populations, a greater proportion of individuals belong to the *Bacteroides* enterotype. For examples, Arumugam *et al.* reported that 63.0%, 13.0%, and 24.0% of 154 Americans were assigned to *Bacteroides*, *Ruminococcus* and *Prevotella* enterotypes, respectively, and that, similarly, most (63.0%) of 85 Danish individuals also were grouped to *Bacteroides* enterotype (Arumugam *et al.*, 2011). Zhou *et al.* also identified the predominance of the *Bacteroides* enterotype (61.2%) in the

HMP healthy cohort (Zhou *et al.*, 2014). The differences in the proportion of *Bacteroides* vs. *Prevotella* enterotypes between Korean and Western populations may be caused by the differences in dietary habits or demographic characteristics of the populations. However, in order to compare more accurately the prevalence of the specific enterotypes across studies, further studies taking methodological differences across studies into account are needed.

With regard to the functional potential of enterotypes, we observed that modules for various vitamin biosynthesis and pentose phosphate pathway were enriched in enterotype 1 (Figure 2.3). This observation is in agreement with a recent study showing that the *Prevotella*-dominated (relatively *Bacteroides*-reduced) gut microbiota of patients with untreated rheumatoid arthritis had a reduced abundance of vitamin metabolism and pentose phosphate pathway modules compared to healthy individuals (Scher *et al.*, 2013). Thus, *Prevotella* are likely to be key microbes leading to differences in the abundances of these functional modules between the enterotypes. Indeed, the enzymes of the pentose phosphate pathway do not exist in *Prevotella* (Shah and Collins, 1990). In line with our observation of the enrichment of various vitamin biosynthesis modules in enterotype 1, *Bacteroides* living in the gut are known to have a high abundance of genes in the KEGG pathway for vitamin and cofactor metabolism (Karlsson *et al.*, 2011). In addition, the enrichment in transport system modules in enterotype 2 implies that the microbes of enterotype 2 may have the high efficiency in the uptake of nutrients from the extracellular environment. It has been reported that at the phylum level, *Firmicutes* have relatively more transport systems than *Bacteroidetes* (Turnbaugh *et al.*, 2009). Several taxa enriched in enterotype 2, such as *Lactobacillus*, *Dorea*, and *Coprococcus*, belong to the *Firmicutes*, which may contribute to the enrichment of transport system modules in enterotype 2.

The MZ twin design has the advantage of matching for age, sex, and genetic background. Another advantage of studying twins is that twins share maternal influences and common environment. Therefore, estimating the concordance rate of the

enterotype can help suggest whether the enterotype is related to host genetics, although in order to better understand host genetic effects on the enterotype, it is required to compare the concordance rate of MZ twins to that of DZ twins. Our finding of a high level of concordance for enterotypes of MZ twins (Table 2.3) indicates that the enterotype might be affected by the host genetics to some extent. However, a part of the MZ twin pairs (27.8%) were found to be enterotype-discordant. This discordance may be partially caused by the fact that all of our subjects were adults (aged 30–48 years at the first time point), and thus the diets and environments of the twin pairs might not have been shared for a long time. In addition, our data showed that the enterotype of each individual was generally stable over time (Figure 2.4). Previous studies on the stability of enterotypes have suggested that neither a 10-day nor 6-month dietary change was sufficient to lead to enterotype switching (Roager *et al.*, 2014, Wu *et al.*, 2011). A recent study with very limited data reported the stability of enterotypes over a long-term period (Rajilić-Stojanović *et al.*, 2012): the enterotypes of all five participating subjects appeared to have changed over a period of 8 or 12 years. To better understand the factors causing the change of an individual's enterotype, larger longitudinal studies with demographic, nutritional, and behavioral data are needed.

Among a broad range of clinical biomarkers, the serum uric acid level showed a significant association with the enterotypes (Figure 2.8). Uric acid is the end product of purine metabolism in humans, and its production is in balance with uric acid disposal under the steady-state condition. Approximately two-thirds of uric acid pool is excreted from the kidney into urine, and the remainder is excreted into gut where uric acid is broken down into allantoin by gut bacteria (intestinal uricolysis) (Sorensen, 1959). We found that 5-hydroxyisourate hydrolase (K07127), which is involved in the conversion of uric acid to allantoin, was enriched in enterotype 1 (data not shown). Thus, differences in the functional capacity to metabolize uric acid between the enterotypes may contribute to differences in serum uric acid levels between the enterotypes. Further studies combining metagenomic, metatranscriptomic, and metabolomic approaches are

needed to determine which enterotype-associated gut microbes affect serum uric acid levels, and how.

Enterotypes were previously reported to be associated with long-term dietary patterns: the "*Bacteroides*" enterotype was correlated with a diet enriched in protein and animal fat, while the "*Prevotella*" enterotype was associated with a carbohydrate-enriched diet (Wu *et al.*, 2011). In the present study, no such association was observed. Instead, we found that long-term intakes of dietary fiber and micronutrients were significantly associated with the enterotypes (Figure 2.9). Our data show that enterotype 2 is overrepresented not only by *Prevotella* but also by several bacteria, such as *Lactobacillus* and *Coprococcus*, which have the ability to ferment dietary fiber (Ahn *et al.*, 2013, Stewart *et al.*, 2009). Indeed, De Filippo *et al.* found that the gut microbiome of rural African children was dominated by saccharolytic bacteria, including *Prevotella* and *Xylanibacter*, compared to that of European children (De Filippo *et al.*, 2010). These differences were attributed to the high dietary fiber intake of the rural children. Even though our study subjects were composed of members of a single ethnic group, we were able to observe the association of dietary fiber intake with enterotypes.

The types of vitamins whose intake levels were found to be associated with the enterotypes were vitamin A, including retinol and carotene, vitamin C, and vitamin E (Figure 2.9B). All of these vitamins have antioxidant properties, and influence immune function by strengthening epithelial barriers and cellular immune responses (Maggini *et al.*, 2007). Moreover, a study with vitamin A-deficient mice revealed that vitamin A can modulate the composition of gut microbiota, and in turn affect the differentiation of pro-inflammatory Th17 cells (Cha *et al.*, 2010). Thus, our data suggest that dietary vitamin intakes may partially contribute to particular configurations of the gut microbiota, and therefore, host immune phenotypes may differ according to the host's enterotype.

We also observed that in enterotype 1, folate intake was lower (Figure 2.9B) and the KEGG module related to folate biosynthesis was significantly increased (Figure 2.3). Folate is a water-soluble B vitamin produced by both gut microbes and plants, and is

involved in various cellular processes including the synthesis of nucleotides and certain amino acids. Therefore, it is likely that enterotype 1 has adapted to compensate for a dietary shortage of this essential nutrient, in agreement with Yatsunenko *et al*.'s observation that the genes related to the riboflavin biosynthesis are more abundant in the gut microbiome of the Malawian and Amerindian infants whose vitamin availability is lower than the U.S. infants (Yatsunenko *et al.*, 2012).

With regard to minerals, higher dietary iron intake was found to be associated with enterotype 2 (Figure 2.9B). Iron is an essential micronutrient for most of the gut microbiota, and the ability to acquire iron differs among bacterial strains. Several studies have reported that iron supplementation leads to compositional changes in the gut microbiota in humans and in animal models (Alexandra *et al.*, 2014, Werner *et al.*, 2011, Zimmermann *et al.*, 2010). Thus, the gut microbial compositions forming the different enterotypes could be influenced by dietary iron intake.

In conclusion, we presented that the gut microbiota of a Korean MZ twin population can be categorized into two enterotypes, each enriched in either *Bacteroides* or *Prevotella*. Our longitudinal samples of MZ twins showed that most co-twins shared their enterotypes, and that the enterotype of an individual remained mostly unchanged over a long-term period. The serum uric acid level and the long-term intakes of nutrients such as dietary fiber, vitamins, and minerals were found to be associated with the subjects' enterotypes.

CHAPTER III.

THE EFFECT OF HERITABILITY AND HOST GENETICS ON THE GUT MICROBIOTA AND METABOLIC SYNDROME

Introduction

Due to recent dramatic changes in human lifestyle, the metabolic syndrome (MetS), which consists of multiple clinical factors leading to high risk of type 2 diabetes and cardiovascular disease (Wilson *et al.*, 2005), has become increasingly prevalent worldwide. It was estimated that approximately 20–30% of adults suffer from MetS in most countries (Grundy, 2008). For example, recent data showed that 34.6% and 31.3% of adults suffer from MetS in the United States and South Korea, respectively (Ford, 2005, Lim *et al.*, 2011). The hallmarks of MetS include abdominal obesity, dyslipidemia, hypertension, and hyperglycemia. These symptoms are closely associated with insulin resistance and chronic low-grade inflammation, which lead to the development of type 2 diabetes and cardiovascular disease (Esposito *et al.*, 2004). Although several factors have been implicated in the development of MetS, including host genetic factors (*e.g.*, single nucleotide polymorphisms (SNPs) in or near lipoprotein lipase (LPL), cholesteryl ester transfer protein (CETP), and apolipoprotein A (APOA)-cluster), dietary habit (*e.g.*, a high fat/high calorie diet), and sedentary lifestyle (*e.g.*, physical inactivity) (Kaur, 2014, Kraja *et al.*, 2011), the pathogenesis of MetS is not yet fully elucidated.

Recently, a number of studies have indicated that the gut microbiota are involved in the energy metabolism of the host, regulating nutrient absorption, gut hormone production, fat storage, or gut permeability (Bäckhed *et al.*, 2004, Cani *et al.*, 2008, Maccarrone *et al.*, 2010, Samuel *et al.*, 2008). In addition, accumulated data have suggested that dysbiosis of the gut microbiota is closely related to the development of obesity and, subsequently, other hallmarks of MetS (Ley *et al.*, 2006, Turnbaugh *et al.*, 2009, Turnbaugh *et al.*, 2006, Vijay-Kumar *et al.*, 2010, Zupancic *et al.*, 2012). These data imply that the gut microbiota might play a critical role in the development of MetS in humans, but few studies have investigated the gut microbiota of individuals in relation to MetS.

The composition of the gut microbiota is modulated by multiple factors, including the age, sex, and diet of the host (Delzenne *et al.*, 2011). In addition, increasing evidence from mouse models has suggested that host genetics also contribute to shaping the gut microbiota (Benson *et al.*, 2010, Kovacs *et al.*, 2011, McKnite *et al.*, 2012, Zhang *et al.*, 2009a). Recently, Goodrich *et al.* found that a number of gut microbial taxa, including Christensenellaceae, were heritable in humans, which strongly suggests that the human gut microbiota can be influenced by host genetics (Goodrich *et al.*, 2014). At the level of specific loci, several studies have shown that variations in several host genes, such as nucleotide-binding oligomerization domain 2 (NOD2) and fucosyltransferase 2 (FUT2), may contribute to alterations in gut microbial community structure, and consequently, affect Crohn's disease (CD) susceptibility in humans (Ateequr *et al.*, 2011, Philipp *et al.*, 2011, Tong *et al.*, 2014). Therefore, it can be hypothesized that MetS, gut microbial composition, and host genotype are interrelated. However, to our knowledge, no studies to date have investigated which specific gut microbes are responsible for MetS, or the degree to which such microbes are related to SNPs in particular host genes.

The *APOA5* gene encodes apolipoprotein A-V, a component of chylomicrons, very-low-density lipoprotein (VLDL), and high-density lipoprotein (HDL) particles (O'Brien *et al.*, 2005). This gene, which is located proximal to the *APOA1/C3/A4* gene cluster on chromosome 11, appears to be involved in the regulation of triglyceride metabolism. Human *APOA5* transgenic mice had greatly reduced plasma triglyceride levels compared to their wild-type counterparts, whereas mice lacking the *APOA5* gene had triglyceride levels fourfold higher than control mice (Pennacchio *et al.*, 2001). In addition, it has been shown in humans that SNPs in the *APOA5* gene are associated with increased levels of plasma triglyceride. The rs651821 SNP, located 3 bp upstream from the start codon of *APOA5*, showed significant associations with increased risk of hypertriglyceridemia or MetS in various populations (Hong *et al.*, 2014, Tan *et al.*, 2012, Yamada *et al.*, 2008). Because *APOA5* is also expressed in the intestine, although the expression level is much lower than that in the liver (Guardiola *et al.*, 2012), we

hypothesized that there would be a relationship between the *APOA5* SNP rs651821 and the gut microbiota, and that such a relationship might be associated with MetS.

Here, we investigated the gut microbiota of 655 Korean twins and their families to identify specific gut microbes significantly associated with MetS status, estimate heritability of the gut microbes, and determine the associations of MetS-related and heritable gut microbes with an *APOA5* genetic variant.

Materials and Methods

Description of study population and their specimens

A total of 655 individuals from participants enrolled in the Healthy Twin Study in Korea (Sung *et al.*, 2006) were included in this study. The study population consisted of 153 monozygotic (MZ) twin pairs (n = 306), 37 dizygotic (DZ) twin pairs (n = 74), and their parents and siblings (n = 275) (Table 3.1). They had not taken any antibiotics or cold medicine 3 months before sampling. Overall, the mean subject age was 47.01 (± 12.19 SD) years, with a range of 20 to 81 years, and 58.47% of the subjects were female. The study protocol was approved by the Institutional Review Board of Samsung Medical Center (IRB file No. 2005-08-113), Busan Paik Hospital (IRB file No. 05-037), and Seoul National University (IRB file No. 144-2011-07-11). Written informed consent was obtained from all participants.

The fecal samples from the participants were obtained at home and immediately frozen in a home freezer. Then, frozen fecal samples were transferred to clinics and stored at –80°C until analysis. All participants answered questionnaires covering lifestyle, medication, and disease history, and underwent clinical tests, biochemical tests, and anthropometrical measurements. Genotyping of the participants was performed using the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Inc., Santa Clara, CA, USA).

Measures of MetS components and definition of MetS

Waist circumference was measured at a level midway between the lowest lateral border of the ribs and the uppermost lateral iliac crest. Blood pressure was measured using a standard manual sphygmomanometer. Blood samples were drawn from an antecubital vein after overnight fasting for at least 8 h. Triglyceride and HDL cholesterol were

assayed using homogeneous and enzymatic methods, respectively. Fasting blood sugar (FBS) was measured using a hexokinase enzymatic method.

MetS was defined following the revised National Cholesterol Education Program Adult Treatment Panel III (NCEP ATP III) criteria (Grundy *et al.*, 2005) with the Korean-specific waist circumference cutoff values for abdominal obesity (Lee *et al.*, 2007). The subjects were considered to have MetS if they had three or more of the following five criteria: 1) waist circumference ≥ 90 cm in males or ≥ 85 cm in females, 2) blood pressure ≥ 130/85 mm Hg, 3) triglycerides ≥ 150 mg/dL (1.7 mmol/L), 4) HDL cholesterol < 40 mg/dL (1.03 mmol/L) in males or < 50 mg/dL (1.3 mmol/L) in females, and 5) FBS ≥ 100 mg/dL (5.6 mmol/L).

Analysis of the gut microbiome

DNA was extracted from the fecal samples using the QIAamp DNA Stool Mini Kit (Qiagen, Inc., Valencia, CA, USA) with the following protocol modifications: Approximately 200 mg of fecal sample were mixed with 1.3-mL ASL buffer in a 0.7-mm garnet bead tube (MO BIO Laboratories, Carlsbad, CA, USA) and vortexed for 30 s. The samples were subsequently heated at 95°C for 10 min and then centrifuged. The supernatant was transferred into a 0.1-mm glass bead tube (MO BIO Laboratories, Carlsbad, CA, USA), homogenized twice by bead beating for 2 min each, and centrifuged. The supernatant was transferred into a new tube, vortexed with an inhibitEX tablet (Qiagen, Inc., Valencia, CA, USA) for 1 min, and incubated for 1 min. The samples were centrifuged, and the supernatant was placed in a new sample tube. The remaining steps of the DNA extraction process were performed by the automated QIAcube instrument (Qiagen, Inc., Valencia, CA, USA) and eluted in 200-μL AE buffer. The extracted DNA samples were stored at –20°C until use.

PCR amplification of the fecal DNA samples was performed with the Illumina-adapted universal primers 515F/806R targeting the V4 region of the 16S rRNA gene

(Caporaso *et al.*, 2011). PCR mixtures (50 μL) contained 35.5-μL PCR water, 5 μL of 10× Takara Ex Taq buffer, 0.1 mM Takara dNTP mix, 0.25 μM of each primer, 0.05 units Ex Taq polymerase (TaKaRa, Shiga, Japan), and 5.0-μL genomic DNA. Reactions were held at 94°C for 3 min to denature the DNA, followed by amplification for 35 cycles at 94°C for 45 s, 50°C for 60 s, and 72°C for 90 s; a final extension at 72°C for 10 min was added to ensure complete amplification. The amplicons were purified using the MO BIO UltraClean PCR Clean-Up Kit (MO BIO Laboratories, Carlsbad, CA, USA) and quantified using the KAPA Library Quantification Kit (KAPA Biosystems, Woburn, MA, USA). The amplicons for each sample were normalized, pooled, and sequenced on the MiSeq platform using a paired-end 2 × 300 bp reagent kit (Illumina, San Diego, CA, USA).

Sequence data were processed using QIIME (v 1.7.0) (Caporaso *et al.*, 2010). Sequences were clustered into operational taxonomic units (OTUs) at 97% identity using an open-reference OTU picking protocol against the 13_5 revision of the Greengenes database. The taxonomy assignments for OTUs were based on the Greengenes reference sequence using the RDP classifier. The resulting tables of OTU counts were subsequently converted to relative abundance tables. Unclassified taxa at each taxonomic level were excluded from the analyses. Phylogenetic investigation of communities by reconstruction of unobserved states (PICRUSt) (Langille *et al.*, 2013) was used to infer putative functional metagenomes from 16S rRNA gene sequence profiles. The relative abundance of each functional pathway was calculated for each sample, and non-microbial functional pathways belonging to the ‘Organismal Systems’ and ‘Human Diseases’ categories were excluded from downstream analysis.

Statistical analysis

Permutational multivariate analysis of variance using distance matrices (Adonis) (Anderson, 2001) and analysis of similarities (ANOSIM) (Clarke, 1993) were performed using the Vegan package in R to identify a significant influence of MetS

status on the gut microbial diversity. These tests were performed using a Bray-Curtis distance matrix calculated from the genus-level relative abundance data. Additionally, two-dimensional nonparametric multi-dimensional scaling (NMDS) plots were generated from a Bray-Curtis distance matrix using the Vegan package in R to visually represent microbiota compositional differences among samples.

Multivariate analysis was performed using multivariate association with linear models (MaAsLin) (Morgan *et al.*, 2012) to identify significant associations of microbial abundances (at all taxonomic levels from kingdom to genus) as well as associations of functional pathway abundances with MetS status, after accounting for confounding variables. In the analysis, age and sex were treated as fixed effects, and MZ twin and family relationships were treated as random variables. Low abundance taxa (the average relative abundance across all the samples < 0.1%) were excluded from MaAsLin analysis. Correlation between the each MetS component and MetS-related microbes was assessed by Spearman's rank correlation test. Associations were considered to be significant with a Benjamini and Hochberg false discovery rate (FDR)-corrected P-value of < 0.25.

Bray-Curtis distances of the genus-level relative abundance profiles between MZ twin pairs, DZ twin pairs, family members, and unrelated individuals were compared by two-sample t-test with 1000 Monte Carlo permutations.

Heritability estimates for each microbial taxon were obtained using a variance component method in Sequential Oligogenic Linkage Analysis Routines (SOLAR) (Almasy and Blangero, 1998). Low-abundance taxa were excluded from this analysis. The relative abundances of the taxa were transformed using the inverse normal command in SOLAR. Heritability estimates (H2r) was calculated after adjustment for covariates including age, sex, and MetS. FDR-corrected P-values of < 0.05 were considered to indicate significance. Heritability estimates for metabolic traits were calculated in the same manner, except that only age and sex were included as covariates. Heritability of MetS itself was analyzed using a threshold model with adjustment for

age and sex in SOLAR.

The associations between host genotype and specific gut microbial taxa were tested using the family-based association tests for quantitative traits (QFAM) module implemented in PLINK (http://pngu.mgh.harvard.edu/purcell/plink/) (Purcell *et al.*, 2007). QFAM performs a simple linear regression of quantitative phenotype on genotype, while accounting for family structure. Because the QFAM do not accept covariates, the gut microbial abundances, which were transformed with the arcsine square root transformation, were regressed on MetS status and/or triglyceride in addition to age and sex, and the standardized residual values were used as quantitative phenotypes in QFAM analysis. A total of 100,000 permutations were carried out to correct for the dependence between related individuals, and the results were reported as pointwise empirical P-value (EMP1). The resulting EMP1 values were adjusted for multiple testing with the FDR correction. The effects of host genotype on each MetS trait were also examined with QFAM. The association of host genotype with MetS status was assessed with the chi-squared test.

Results

Association of gut microbial community structure and MetS status

Among the 655 individual in our study, the subjects who fulfilled MetS criteria represented 18.47% of the total, 12.75% of MZ twins, 13.51% of DZ twins, and 26.18% of family members (Table 3.1). To assess the strength of relationships between microbial community diversity and MetS phenotype, we first performed two different nonparametric analyses for multivariate data, including Adonis and ANOSIM statistical tests, on the Bray-Curtis distance matrix for the genus-level relative abundance profiles. The analyses showed that the gut microbial communities of MetS individuals were significantly different from those of healthy individuals (Adonis: $P = 0.034$, ANOSIM: $P = 0.028$), although the gut microbiota from these two groups were not clearly separated on a two-dimensional NMDS plot (Figure 3.1).

To evaluate the associations of gut microbial taxa with MetS status, with adjustments for age, sex, and correlation structures from identical twin and family relationships, we conducted MaAsLin analysis (Morgan *et al.*, 2012) (Figure 3.2A). The results indicated that *Lactobacillus*, *Methanobrevibacter*, and the *Sutterella* were markedly increased in MetS individuals, while *Parabacteroides*, *Odoribacter*, *Akkermansia*, *Bifidobacterium*, and the Christensenellaceae were significantly overrepresented in healthy individuals.

For the taxa identified by MaAsLin, we confirmed correlations between the gut microbial abundances and each MetS component, including waist circumference, triglycerides, FBS, blood pressure, and HDL cholesterol, using Spearman's rank correlation test. We observed a tendency in which the taxa enriched in the MetS individuals were positively correlated with waist circumference, triglycerides, blood pressure, or FBS, but negatively correlated with HDL cholesterol, while the opposite trend was observed for the taxa enriched in healthy individuals, although there were several exceptions (Figure 3.2B). For example, *Lactobacillus*, enriched in the MetS

individuals, was positively correlated with FBS and waist circumference, but negatively correlated with HDL cholesterol. In contrast, *Odoribacter* and *Rikenellaceae*, overrepresented in healthy individuals, were positively correlated with HDL cholesterol, and negatively correlated with all other MetS components. A similar tendency was observed even after adjustment for age, sex, and family structure (Figure 3.2C).

To further investigate the gut microbial functions associated with MetS status, we used PICRUSt (Langille *et al.*, 2013) to predict functional metagenomes from 16S rRNA gene sequences and their reference genomes. Functional pathways that were differentially abundant between healthy and MetS individuals were identified, with adjustments for age, sex, and family structure, using MaAsLin (Morgan *et al.*, 2012). We found that cysteine and methionine metabolism, nicotinate and nicotinamide metabolism, vitamin B6 metabolism, and peptidoglycan biosynthesis pathways were overrepresented in MetS individuals, whereas the pathways associated with carbohydrate, lipid, and energy metabolism (*e.g.*, pyruvate, glycerophospholipid, and methane metabolism), cell motility (*e.g.*, bacteria chemotaxis, flagellar assembly, and bacterial motility proteins), and genetic information processing (e.g., RNA transport and the sulfur-relay system) were enriched in healthy individuals (Figure 3.3). Thus, we concluded that the gut microbiota of MetS individuals are compositionally and functionally distinct from the gut microbiota of healthy individuals.

Table 3.1 Characteristics of participants in this study.

	n (%) or mean ± SD			
	Overall	**MZ twins**	**DZ twins**	**Nontwin**
	n=655	**n=306**	**n=74**	**n=275**
Females	383 (58.47)	184 (60.13)	36 (48.65)	163 (59.27)
Age (yr)	47.01 ± 12.19	41.83 ± 7.99	41.23 ± 7.57	54.32 ± 13.26
Waist ≥ 90 (M) or 85 (F) cm	156 (23.82)	53 (17.32)	9 (12.16)	94 (34.18)
BP ≥ 130/85 mm Hg	104 (15.88)	32 (10.46)	12 (16.22)	60 (21.82)
Triglyceride ≥ 150 mg/dL	164 (25.04)	63 (20.59)	20 (27.03)	81 (29.45)
HDL < 40 mg/dL (M) or 50 mg/dL (F)	296 (45.19)	122 (39.87)	29 (39.19)	145 (52.73)
FBS ≥ 100 mg/dL	158 (24.12)	58 (18.95)	12 (16.22)	88 (32)
MetS*	121 (18.47)	39 (12.75)	10 (13.51)	72 (26.18)

M: males; F: females; Waist: waist circumference; BP: blood pressure; HDL: high-density lipoprotein cholesterol; FBS: fasting blood sugar; MetS: metabolic syndrome; MZ: monozygotic; DZ: dizygotic; Nontwin: parents or siblings of twin pairs

*Abnormal values for at least three of the following: Waist, BP, Triglyceride, HDL, and FBS

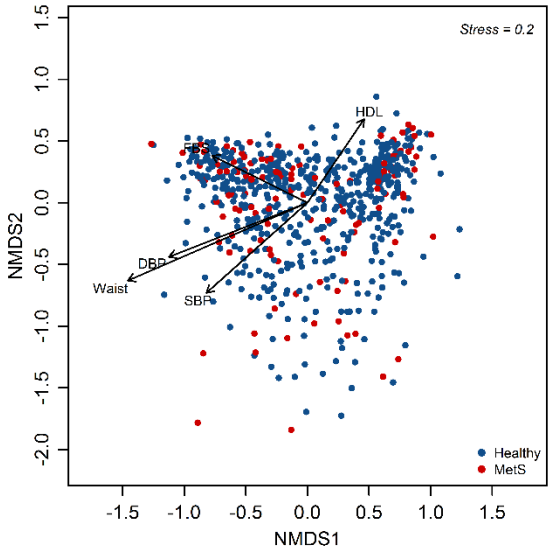
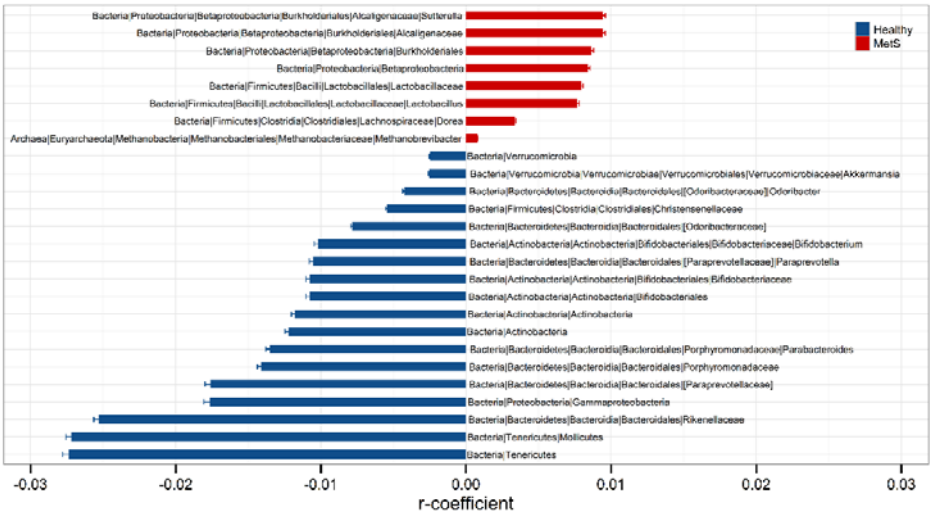
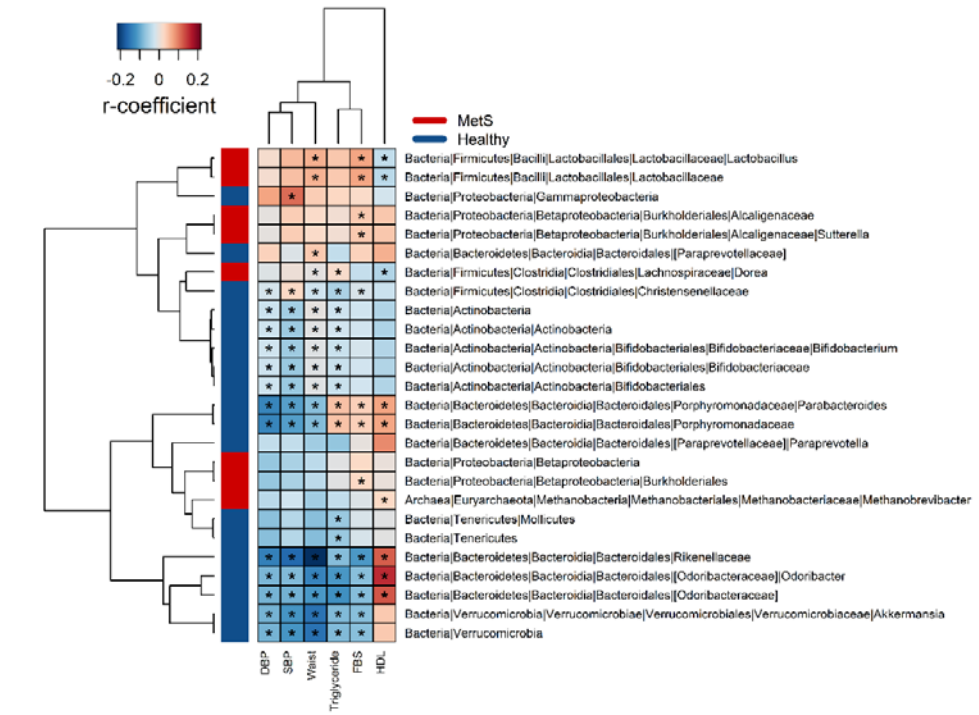


Figure 3.1 Nonparametric multi-dimensional scaling (NMDS) plots of gut microbiota. Bray-Curtis distance matrix calculated from the genus-level relative abundance data were used in this NMDS analysis. Red and blue dots represent MetS individuals and healthy individuals, respectively. Arrows indicate MetS components that had significant correlations with the ordination. (DBP: diastolic blood pressure; SBP: systolic blood pressure; FBS: fasting blood sugar; HDL: high-density lipoprotein cholesterol; Waist: waist circumference).

(A)


Bacteria|Proteobacteria|Betaproteobacteria|Burkholderiales|Alcaligenaceae|Sutterella
Bacteria|Proteobacteria|Betaproteobacteria|Burkholderiales|Alcaligenaceae
Bacteria|Proteobacteria|Betaproteobacteria|Burkholderiales
Bacteria|Proteobacteria|Betaproteobacteria
Bacteria|Firmicutes|Bacilli|Lactobacillales|Lactobacillaceae
Bacteria|Firmicutes|Bacilli|Lactobacillales|Lactobacillaceae|Lactobacillus
Bacteria|Firmicutes|Clostridia|Clostridiales|Lachnospiraceae|Dorea
Archaea|Euryarchaeota|Methanobacteria|Methanobacteriales|Methanobacteriaceae|Methanobrevibacter
Bacteria|Verrucomicrobia
Bacteria|Verrucomicrobia|Verrucomicrobiae|Verrucomicrobiales|Verrucomicrobiaceae|Akkermansia
Bacteria|Bacteroidetes|Bacteroidia|Bacteroidales|[Odoribacteraceae]|Odoribacter
Bacteria|Firmicutes|Clostridia|Clostridiales|Christensenellaceae
Bacteria|Bacteroidetes|Bacteroidia|Bacteroidales|[Odoribacteraceae]
Bacteria|Actinobacteria|Actinobacteria|Bifidobacteriales|Bifidobacteriaceae|Bifidobacterium
Bacteria|Bacteroidetes|Bacteroidia|Bacteroidales|[Paraprevotellaceae]|Paraprevotella
Bacteria|Actinobacteria|Actinobacteria|Bifidobacteriales|Bifidobacteriaceae
Bacteria|Actinobacteria|Actinobacteria|Bifidobacteriales
Bacteria|Actinobacteria|Actinobacteria
Bacteria|Actinobacteria
Bacteria|Bacteroidetes|Bacteroidia|Bacteroidales|Porphyromonadaceae|Parabacteroides
Bacteria|Bacteroidetes|Bacteroidia|Bacteroidales|Porphyromonadaceae
Bacteria|Bacteroidetes|Bacteroidia|Bacteroidales|[Paraprevotellaceae]
Bacteria|Proteobacteria|Gammaproteobacteria
Bacteria|Bacteroidetes|Bacteroidia|Bacteroidales|Rikenellaceae
Bacteria|Tenericutes|Mollicutes
Bacteria|Tenericutes
Healthy
MetS
-0.03
-0.02
-0.01
0.00
0.01
0.02
0.03
r-coefficient

(B)


-0.2
0
0.2
r-coefficient
MetS
Healthy
Bacteria|Firmicutes|Bacilli|Lactobacillales|Lactobacillaceae|Lactobacillus
Bacteria|Firmicutes|Bacilli|Lactobacillales|Lactobacillaceae
Bacteria|Proteobacteria|Gammaproteobacteria
Bacteria|Proteobacteria|Betaproteobacteria|Burkholderiales|Alcaligenaceae
Bacteria|Proteobacteria|Betaproteobacteria|Burkholderiales|Alcaligenaceae|Sutterella
Bacteria|Bacteroidetes|Bacteroidia|Bacteroidales|[Paraprevotellaceae]
Bacteria|Firmicutes|Clostridia|Clostridiales|Lachnospiraceae|Dorea
Bacteria|Firmicutes|Clostridia|Clostridiales|Christensenellaceae
Bacteria|Actinobacteria
Bacteria|Actinobacteria|Actinobacteria
Bacteria|Actinobacteria|Actinobacteria|Bifidobacteriales|Bifidobacteriaceae|Bifidobacterium
Bacteria|Actinobacteria|Actinobacteria|Bifidobacteriales|Bifidobacteriaceae
Bacteria|Actinobacteria|Actinobacteria|Bifidobacteriales
Bacteria|Bacteroidetes|Bacteroidia|Bacteroidales|Porphyromonadaceae|Parabacteroides
Bacteria|Bacteroidetes|Bacteroidia|Bacteroidales|Porphyromonadaceae
Bacteria|Bacteroidetes|Bacteroidia|Bacteroidales|[Paraprevotellaceae]|Paraprevotella
Bacteria|Proteobacteria|Betaproteobacteria
Bacteria|Proteobacteria|Betaproteobacteria|Burkholderiales
Archaea|Euryarchaeota|Methanobacteria|Methanobacteriales|Methanobacteriaceae|Methanobrevibacter
Bacteria|Tenericutes|Mollicutes
Bacteria|Tenericutes
Bacteria|Bacteroidetes|Bacteroidia|Bacteroidales|Rikenellaceae
Bacteria|Bacteroidetes|Bacteroidia|Bacteroidales|[Odoribacteraceae]|Odoribacter
Bacteria|Bacteroidetes|Bacteroidia|Bacteroidales|[Odoribacteraceae]
Bacteria|Verrucomicrobia|Verrucomicrobiae|Verrucomicrobiales|Verrucomicrobiaceae|Akkermansia
Bacteria|Verrucomicrobia
DBP
SBP
Waist
Triglyceride
FBS
HDL

(C)

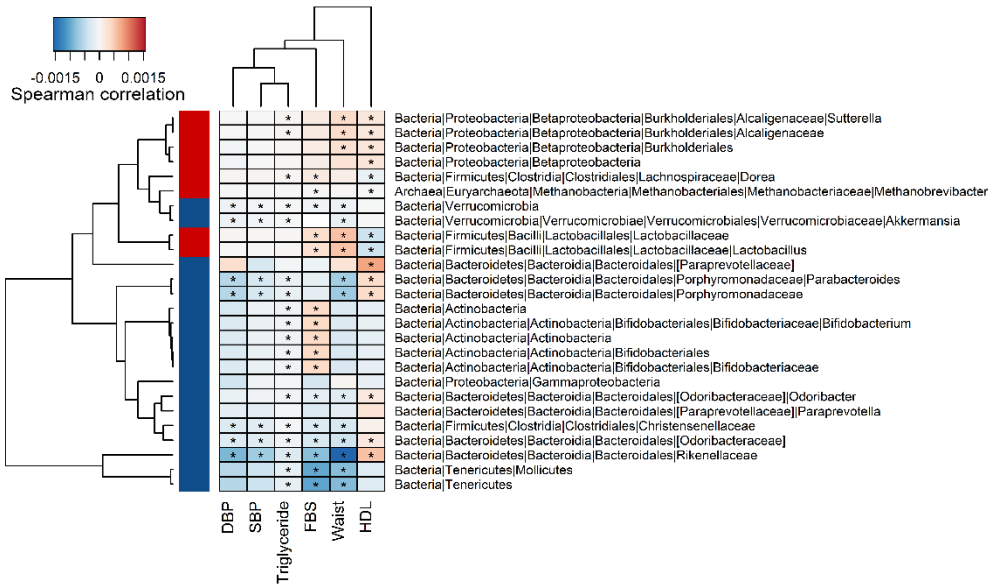


Figure 3.2 Significant differences in the composition of the gut microbiota between the healthy and MetS groups. (A) Bar plot of γ-coefficients from multivariate analysis assessing associations between microbial taxa and MetS status, with adjustments for age, sex, and family structures. Positive (red bar) and negative (blue bar) coefficient values represent taxa enriched in the MetS group and healthy group, respectively. The error bar represents the standard error. B-C: Heat maps of (B) Spearman's rank correlation coefficients and (C) γ-coefficients from multivariate analysis between MetS-associated gut microbial taxa and each MetS component (DBP: diastolic blood pressure; SBP: systolic blood pressure; Waist: waist circumference; Triglyceride; FBS: fasting blood sugar; HDL: high-density lipoprotein cholesterol). Red and blue in the row-side color bar on left-hand side indicate that the taxa were significantly enriched in MetS individuals and healthy individuals, respectively, in MaAsLin analysis. The asterisk indicate significant associations at FDR-corrected P-values of < 0.25.

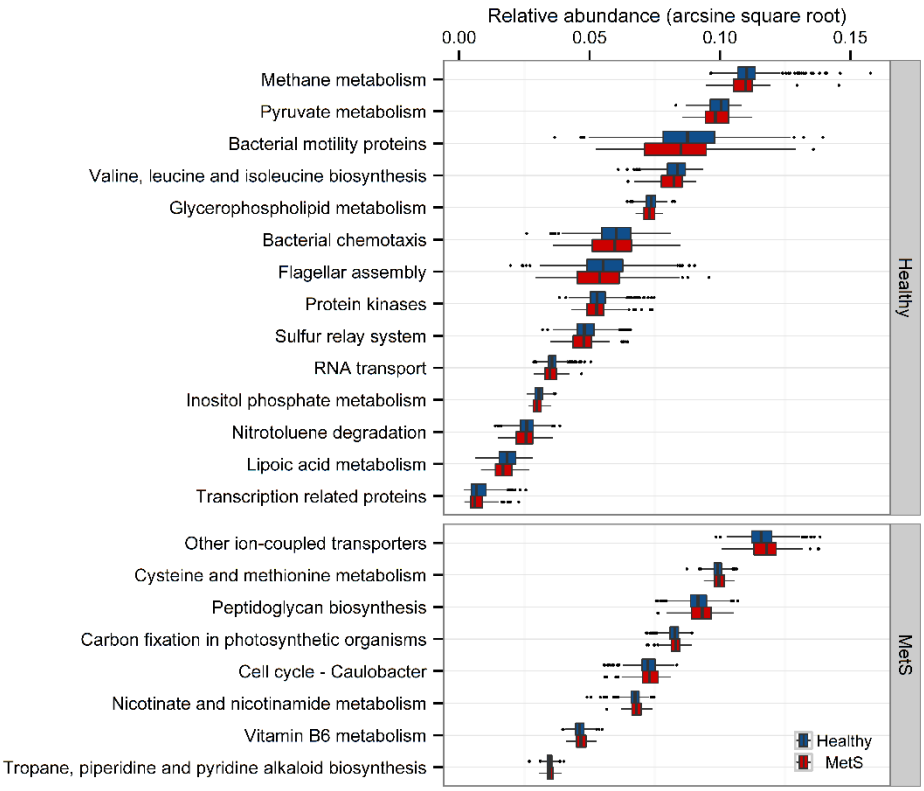


Figure 3.3 Significant associations between inferred gut microbial functions and MetS status. Box plots of relative abundances (arcsine square root transformed values) of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways that were significantly differentially abundant between healthy and MetS individuals. The significant associations were assessed by multivariate analysis, with adjustments for age, sex, and family structures.

Heritability of MetS-associated gut microbial populations

It is known that MetS is significantly influenced by host genetics (Bellia *et al.*, 2009, Lin *et al.*, 2005, Sung *et al.*, 2009). We evaluated the heritability of MetS and its individual traits with adjustments for age and sex. Consistent with a previous study conducted with Korean twins (Sung *et al.*, 2009), MetS exhibited high heritability (70.3%). In addition, all of the MetS components exhibited significant heritability, ranging from 23.9% for diastolic blood pressure (DBP) to 73.5% for waist circumference (Figure 3.4A).

To assess the influences of host genetics on the gut microbiota, we measured the Bray-Curtis distances between the gut microbial communities of MZ twin pairs, DZ twin pairs, family members, and unrelated individuals, based on the genus-level relative abundance profiles. The results indicated that the overall gut microbial community structures were not more similar between MZ twin pairs than between DZ twin pairs ($P = 0.332$; two-sample t-test, 1000 Monte Carlo permutations), although MZ twin pairs had more similar gut microbial communities compared to those from members of a family or unrelated individuals ($P = 0.022$ and $P < 2.2e\text{-}16$) (Figure 3.4B). We also observed the similar results from the Bray-Curtis, unweighted UniFrac, and weighted UniFrac distances that were calculated with a rarefied OTU table (Figure 3.5).

Because MetS is highly heritable (Figure 3.4A) and the gut microbiota is closely related to MetS status (Figure 3.2), we calculated H2r for each taxa after adjusting not only for age and sex, but also for MetS, to assess the genetic influence on the abundance of each taxa, independently of MetS status. Among the 85 tested microbial taxa from genus level to kingdom level with greater than 0.1% average relative abundance, 50 taxa (58.8%) were significantly heritable (FDR-corrected $P < 0.05$), with heritability estimates ranging between 13.1% and 45.7% (Appendix A). Focusing on the taxa found to be significantly associated with MetS status by MaAsLin, we determined that 17 (65.4%) of the 26 taxa were heritable (Figure 3.4C). Among the taxa enriched in the MetS group, *Methanobrevibacter* (H2r ± SE; 20.9 ± 5.9%) and *Lactobacillus* (14.8 ±

5.9%) were significantly heritable. On the other hand, among the taxa enriched in the healthy group, the Actinobacteria phylum (45.7 ± 5.5%), *Bifidobacterium* (37.4 ± 9.4%), Christensenellaceae (30.6 ± 5.9%), and *Odoribacter* (19.2 ± 5.8%) exhibited significant heritability. These results indicate that host genetic factors greatly influence the relative abundances of specific gut microbes that are related to MetS.

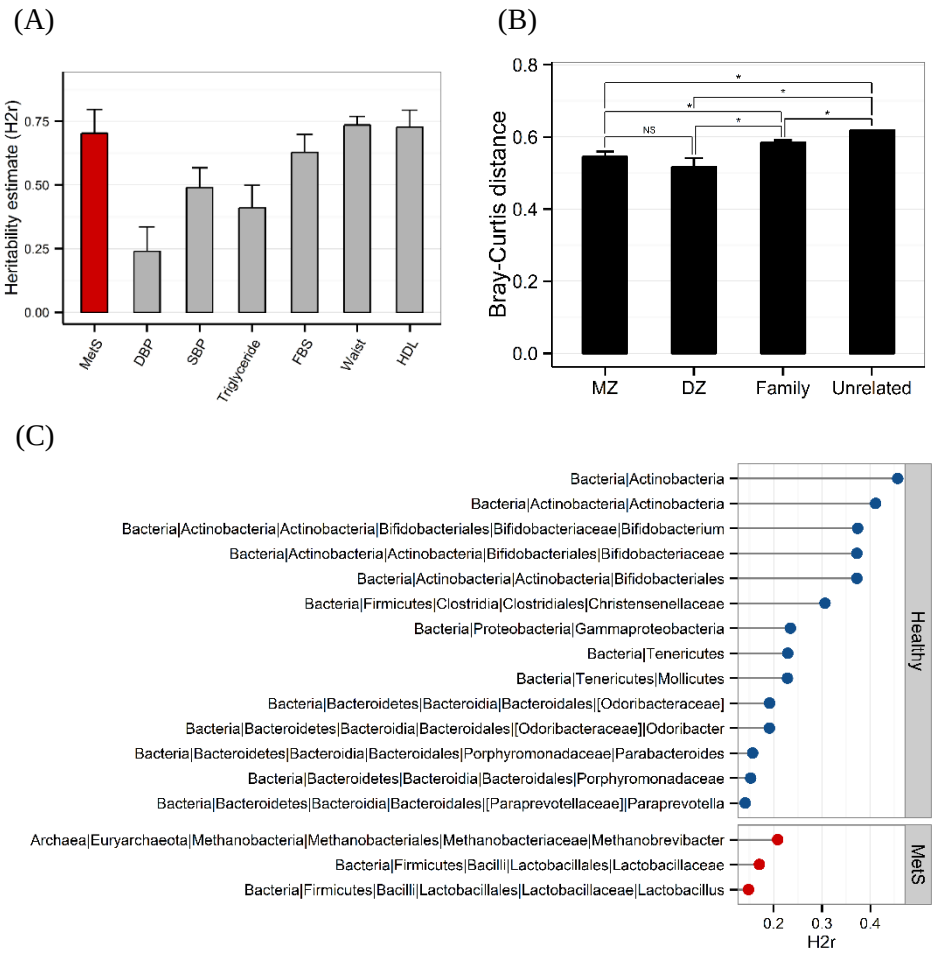


Figure 3.4 Heritability of MetS phenotype and gut microbiota. (A) Bar graph of the heritability estimates (H2r) of MetS status as well as each MetS component (DBP: diastolic blood pressure; SBP: systolic blood pressure; Triglyceride; FBS: fasting blood sugar; Waist: waist circumference; HDL: high-density lipoprotein cholesterol). The error bar represents the standard error of the estimates. (B) Bar graph of Bray-Curtis distances of the genus-level relative abundance profiles between monozygotic (MZ) twin pairs, dizygotic (DZ) twin pairs, family members, and unrelated individuals (mean ± standard error of the mean; *P < 0.05 for two-sample t-test with 1000 Monte Carlo

permutations). (C) Dot plot of the heritability estimates (H2r) of the gut microbial taxa. Red dots and blue dots denote the heritability estimates of taxa that were significantly enriched in MetS individuals and healthy individuals, respectively, in MaAsLin analysis. Heritability values were obtained for inverse-normal-transformed relative abundances of the gut microbial taxa after adjustments for age, sex, and MetS status. Only taxa with FDR-corrected P-values < 0.05 are shown.

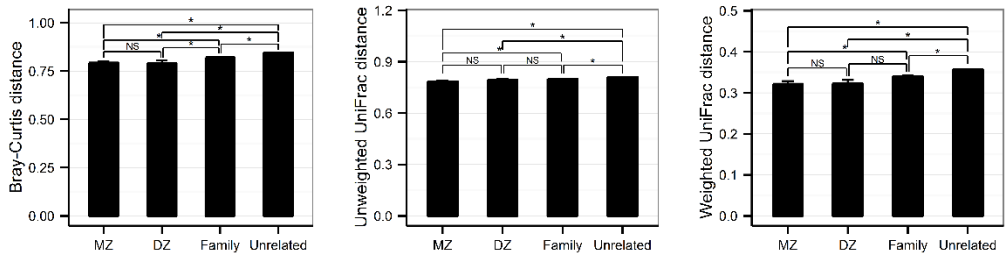


Figure 3.5 Differences in the gut microbiota between family members. Bar graph of Bray-Curtis, unweighted UniFrac, and weighted UniFrac distances between the gut microbial communities of monozygotic (MZ) twin pairs, dizygotic (DZ) twin pairs, family members, and unrelated individuals (mean ± standard error of the mean; *P < 0.05 for two-sample t-test with 1000 Monte Carlo permutations). These distances were calculated with a rarefied OTU table containing 8,000 sequences per sample.

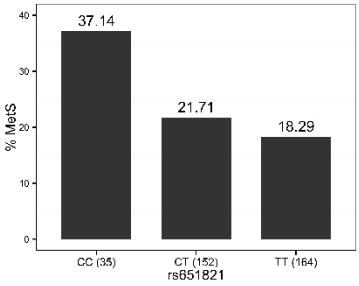
Association of *APOA5* SNP rs651821 with MetS-related gut microbial taxa

To further investigate whether a specific host gene can influence abundances of specific gut microbial taxa, we performed a targeted analysis of *APOA5* SNP rs651821. Because this SNP has been previously associated with triglyceride levels and MetS (Hong *et al.*, 2014, Tan *et al.*, 2012, Yamada *et al.*, 2008), we first tested whether such a relationship existed in this study population with total of 351 individuals (275 healthy and 76 MetS) whose rs651821 genotype data were available (minor allele frequency = 33.1%). We observed that the prevalence of MetS was higher in carriers of the minor C allele than in non-carriers (P = 0.0488; chi-square test) (Figure 3.6A). Moreover, we found that each minor copy of rs651821 was significantly associated with a 24.65 mg/dL increase in triglyceride (EMP1 = 0.0011) using the QFAM procedure in PLINK (Purcell *et al.*, 2007). There were trends of greater increases in DBP, SBP, FBS, and waist circumference and of greater reduction in HDL cholesterol in carriers of the minor allele, although none of these trends were statistically significant (Figure 3.6B). These findings confirmed that minor alleles of *APOA5* SNP rs651821 were closely linked to increased risk of high triglyceride levels as well as MetS in our cohort.

Next, we further tested for associations between *APOA5* SNP rs651821 and each of 17 MetS-related and heritable gut microbial taxa (listed in Figure 3.4C) using the QFAM module. After adjustment for age and sex, five significant associations were detected between SNP rs651821 and gut microbial taxa (FDR-corrected P < 0.05). Each additional copy of the minor C allele at *APOA5* SNP rs651821 decreased the abundances of the Actinobacteria phylum, the Actinobacteria class, Bifidobacteriales, Bifidobacteriaceae, and *Bifidobacterium* (Table 3.2). Notably, although these taxa were found to be significantly related to MetS status (Figure 3.2), SNP rs651821 remained associated with the abundances of these taxa even after adjusting for MetS status, age, and sex (Table 3.3 and Figure 3.7). The associations between these taxa and SNP rs651821 were also observed even after adjustment for triglyceride additionally. Collectively, these data demonstrate that *APOA5* SNP rs651821 might be an important

host genetic factor in determining the abundances of the health-promoting *Bifidobacterium* bacteria and the taxa to which this genus belongs.

(A)



(B)

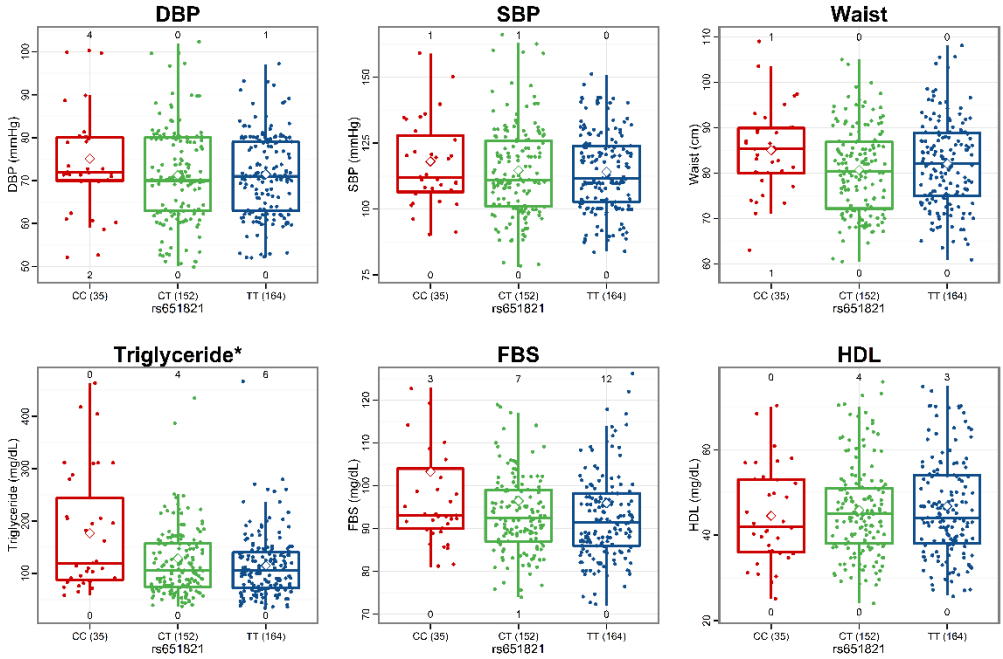


Figure 3.6 Associations of *APOA5* SNP rs651821 with MetS. (A) Prevalence of MetS according to *APOA5* SNP rs651821. Chi-square test: P-value = 0.0488 in the comparison of the prevalence of MetS among the genotypes. (B) Box plots of each MetS component according to *APOA5* SNP rs651821. Triglyceride increased significantly in carriers of the minor allele (EMP1 = 0.0011 in the QFAM procedure). The box plots

indicate the median (horizontal solid line), mean (diamond), interquartile range (IQR) between the first and third quartiles (box), minimum and maximum values excluding outliers (whiskers), and the number of outliers.

Table 3.2 Significant associations of *APOA5* SNP rs651821 with gut microbial taxa using the QFAM procedure in PLINK (adjusted for age and sex).

Taxon	BETA	EMP1	FDR.EMP1
p__Actinobacteria	-0.2713	0.0020	0.0102
p__Actinobacteria|g__Bifidobacterium	-0.2632	0.0026	0.0102
p__Actinobacteria|f__Bifidobacteriaceae	-0.2586	0.0029	0.0102
p__Actinobacteria|o__Bifidobacteriales	-0.2586	0.0029	0.0102
p__Actinobacteria|c__Actinobacteria	-0.2583	0.0030	0.0102
p__Bacteroidetes|g__Paraprevotella	0.0961	0.2058	0.5759
p__Tenericutes	-0.0907	0.3030	0.5759
p__Tenericutes|c__Mollicutes	-0.0901	0.3090	0.5759
p__Bacteroidetes|g__Odoribacter	0.0802	0.3420	0.5759
p__Proteobacteria|c__Gammaproteobacteria	-0.0748	0.3961	0.5759
p__Bacteroidetes|g__Parabacteroides	0.0739	0.4022	0.5759
p__Bacteroidetes|f__Porphyromonadaceae	0.0732	0.4065	0.5759
p__Firmicutes|f__Lactobacillaceae	-0.0445	0.6479	0.7709
p__Firmicutes|f__Christensenellaceae	0.0428	0.6751	0.7709
p__Firmicutes|g__Lactobacillus	-0.0410	0.6802	0.7709
p__Euryarchaeota|g__Methanobrevibacter	0.0286	0.7331	0.7789
p__Bacteroidetes|f__[Odoribacteraceae]	0.0055	0.9462	0.9462

The standardized residuals of the arcsine square root transformed microbial abundances regressed against age and sex were used as quantitative phenotypes in this analysis.

BETA: the effect size (regression coefficient); EMP1: pointwise empirical p-value; FDR.EMP1: FDR-corrected EMP1

Table 3.3 Significant associations of *APOA5* SNP rs651821 with gut microbial taxa using the QFAM procedure in PLINK (adjusted for MetS status, age, and sex).

Taxon	BETA	EMP1	FDR.EMP1
p__Actinobacteria	-0.2541	0.0037	0.0160
p__Actinobacteria|g__Bifidobacterium	-0.2477	0.0038	0.0160
p__Actinobacteria|f__Bifidobacteriaceae	-0.2429	0.0045	0.0160
p__Actinobacteria|c__Actinobacteria	-0.2418	0.0047	0.0160
p__Actinobacteria|o__Bifidobacteriales	-0.2429	0.0047	0.0160
p__Bacteroidetes|g__Paraprevotella	0.1108	0.1441	0.4083
p__Bacteroidetes|g__Parabacteroides	0.0948	0.2795	0.5527
p__Bacteroidetes|f__Porphyromonadaceae	0.0943	0.2806	0.5527
p__Bacteroidetes|g__Odoribacter	0.0899	0.2926	0.5527
p__Tenericutes	-0.0653	0.4605	0.7194
p__Tenericutes|c__Mollicutes	-0.0649	0.4655	0.7194
p__Proteobacteria|c__Gammaproteobacteria	-0.0566	0.5219	0.7394
p__Firmicutes|f__Christensenellaceae	0.0550	0.5903	0.7396
p__Euryarchaeota|g__Methanobrevibacter	0.0421	0.6091	0.7396
p__Firmicutes|f__Lactobacillaceae	-0.0412	0.6694	0.7490
p__Firmicutes|g__Lactobacillus	-0.0375	0.7049	0.7490
p__Bacteroidetes|f__[Odoribacteraceae]	0.0226	0.7895	0.7895

The standardized residuals of the arcsine square root transformed microbial abundances regressed against age, sex, and MetS status were used as quantitative phenotypes in this analysis.

BETA: the effect size (regression coefficient); EMP1: pointwise empirical p-value; FDR.EMP1: FDR-corrected EMP1

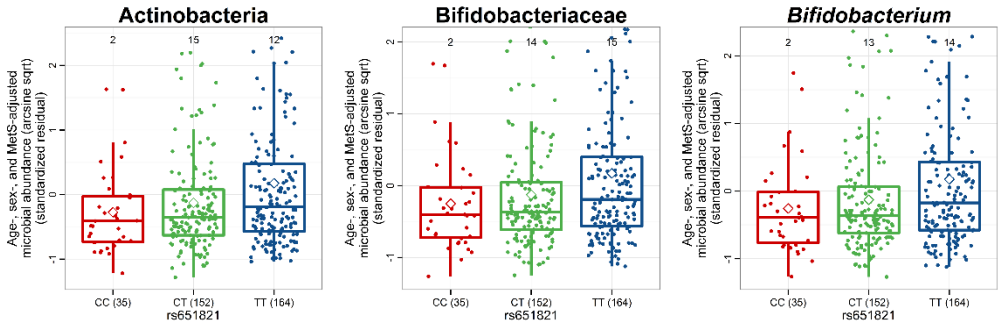


Figure 3.7 Associations of *APOA5* SNP rs651821 with specific members of the gut microbiota. Box plots of the standardized residuals of the arcsine square root transformed microbial abundances regressed against age, sex, and MetS according to *APOA5* SNP rs651821. The box plots indicate the median (horizontal solid line), mean (diamond), the interquartile range (IQR) between the first and third quartiles (box), minimum and maximum values excluding outliers (whiskers), and the number of outliers.

Discussion

Metabolic disorders, including MetS, obesity, and type 2 diabetes, are complex diseases that result from interactions between the host's intrinsic characteristics and environmental factors. Although it is now widely accepted that the gut microbiota is a key environmental factor for modulating host metabolism, only a few studies have pointed to specific gut microbes in relation to MetS, and our understanding of the links between host genetic factors, gut microbiota, and MetS in humans is still quite limited. In this study, we demonstrated that host genetics strongly influences the abundance of specific gut microbial taxa related to MetS. In particular, the *Bifidobacterium* genus is strongly associated with host genetics and its abundance is significantly reduced with each additional copy of the minor allele at *APOA5* SNP rs651821, leading to an increased risk of MetS.

We identified a specific set of gut microbial taxa that may have beneficial or harmful impacts on host metabolism related to MetS (Figure 3.2). A number of studies has focused on microbial associations with obesity and type 2 diabetes in humans, and the results of these studies have been well summarized in reviews (Sanz *et al.*, 2014, Tagliabue and Elli, 2013). Although the results are not fully consistent (Appendix B), we observed that the patterns of gut microbial alterations in MetS individuals were generally similar to those of patients with obesity or type 2 diabetes (Sanz *et al.*, 2014, Tagliabue and Elli, 2013). For example, the reduction in *Bifidobacterium* and the expansion of Betaproteobacteria and Alcaligenaceae that we observed in MetS individuals have also been reported to be associated with obesity or type 2 diabetes (Larsen *et al.*, 2010, Wu *et al.*, 2010, Zhang *et al.*, 2009b). According to a meta-analysis of obesity-associated gut microbiota alterations (Angelakis *et al.*, 2012), *Bifidobacterium* has been shown to have a consistent anti-obesity effect in humans. We also found that *Akkermansia*, whose anti-obesity functions were demonstrated in diet-

induced obese mice (Everard *et al.*, 2013), was enriched in the healthy individuals, although there have been conflicting results on the association between *Akkermansia* abundance and type 2 diabetes in humans (Qin *et al.*, 2012, Zhang *et al.*, 2013). Additionally, we newly identified several gut microbial taxa that may be responsible for the improvement of host metabolism, such as the Porphyromonadaceae family, including the genera *Parabacteroides*, the Rikenellaceae family, and the genus *Odoribacter*. It is noteworthy that the relative abundances of these taxa were not only enriched in the healthy individuals but also associated with improving at least one of the MetS features (Figure 3.2B). Generally, *Lactobacillus* is considered to be beneficial to human health. However, we found an increase in this genus in the gut microbiota of the MetS individuals. Indeed, inconsistent results have been reported with respect to the association between *Lactobacillus* abundance and obesity (Armougom *et al.*, 2009, Balamurugana *et al.*, 2010, Kadooka *et al.*, 2010, Million *et al.*, 2011, Million *et al.*, 2013). Seventeen species of *Lactobacillus* are associated with the human gut (Walter, 2008), and different functional capacities to attenuate obesity complications were *Lactobacillus* strains specific (Wang *et al.*, 2015). These characteristics of *Lactobacillus* spp. may have caused the considerable controversy over the effect of *Lactobacillus* on the control of adiposity. Taken together, although the mechanisms of interaction between these specific gut microbes and the host metabolism remain to be elucidated, our findings suggest that specific modulations of the gut microbiota might be an effective strategy for preventing or treating MetS.

Our data show that host genetics can influence the abundances of specific taxa. The overall gut microbial community structures of MZ twin pairs had an almost identical degree of similarity to that of DZ twin pairs, as shown in previous studies (Turnbaugh *et al.*, 2009, Yatsunenko *et al.*, 2012). This may be due to the influences of early life environmental exposures or to limited statistical power, as recently pointed out by Goodrich *et al.* (Goodrich *et al.*, 2014). In spite of these results, we could not exclude the possibility that specific taxa are influenced by host genetic factors. By applying

variance components methods, we found that more than half of the tested taxa were significantly heritable, and the higher taxonomic levels that encompass the *Bifidobacterium* (Actinobacteria, Actinomycetales, and Bifidobacteriaceae) had the highest heritability estimates among the taxa related to MetS status (Figure 3.4C). Consistent with our results, the heritability of Bifidobacteriaceae was also reported in the twins in UK (Goodrich *et al.*, 2014). Because *Bifidobacterium* dominates the gut microbiota of breast-fed infants and decreases with increasing age (Yatsunenko *et al.*, 2012), it can be expected that the abundance of this genus in adult twins could be affected by an early shared environment. However, surprisingly, no significant influence of the common environments was detected in our study (C2: 0.0198, FDR-corrected P = 0.4620) (Appendix A). Goodrich *et al*. also demonstrated that the Christensenellaceae family was the most heritable taxon in the UK twins, and this family was associated with a lean host phenotype (Goodrich *et al.*, 2014). Consistent with this previous study, we observed the high heritability of the Christensenellaceae family, which was enriched in the healthy individuals in our Korean population. Therefore, the families Bifidobacteriaceae and Christensenellaceae are likely to play important roles in improving host metabolism, interacting with host genetics regardless of ethnicity, diet, or geographic region. We note again that we estimated the heritability of each taxon after removing the effects of MetS status as well as age and sex. Therefore, our data suggest that host genetic factors can influence the development of MetS by determining the abundances of specific gut microbial taxa that are associated with MetS status.

Through a targeted analysis of the association of *APOA5* SNP rs651821 with MetS-related and heritable taxa, we observed that minor alleles of this host genetic variant were significantly associated with decreased abundances of *Bifidobacterium* and its parent taxa, which were simultaneously the health-associated and most heritable taxa (Figure 3.7). This finding suggests that the *APOA5* variant may contribute to the compositional change of the MetS-related gut microbiota. The *APOA5* gene is predominantly expressed in the liver, but it is also produced at a much lower level in the

intestine (Guardiola *et al.*, 2012). The role of *APOA5* as an important regulator of triglyceride metabolism has been demonstrated extensively in mice and human studies, which have mainly focused on liver and plasma (Pennacchio *et al.*, 2001, Ribalta *et al.*, 2002, Schaap *et al.*, 2004, van der Vliet *et al.*, 2002). Recently, a novel role of *APOA5* in the gut was reported by a study of *APOA5*-deficient mice: *APOA5* plays a role in the lymphatic transport of dietary lipids by modulating the production of chylomicrons by the gut (Zhang *et al.*, 2015). Based on our results and the results of this recent study, it is likely that the gut environment of carriers of the minor allele of *APOA5* SNP rs651821 may be different from that of noncarriers in terms of lipid metabolism, which may lead to unfavorable conditions for the growth of *Bifidobacterium*. Interestingly, in a recent CD patient-based study combining host transcriptomic and microbial profiling, the authors found that *APOA1*, *APOA4*, and *APOC3* gene expressions within the CD ileum were downregulated and associated with abundances of specific Firmicutes and Bacteroidetes taxa (Haberman *et al.*, 2014). Although they did not report the expression of the *APOA5* gene, our results are in line with their findings in that apolipoproteins in the *APOA1/C3/A4/A5* gene cluster are associated with specific microbial taxa, and this association would be important for gut microbiota-related diseases. Our study demonstrated that *APOA5* gene determines the abundance of *Bifidobacetrium* independent of both MetS and triglyceride levels.

This study also has limitations that should be noted. First, this was a candidate gene study. The host genetic influence on the gut microbial composition could be derived from the effects of various and multiple genetic variants such as SNPs, copy-number variations, or mutations. To gain a thorough understanding of the associations between host genetic variants and microbial taxa, it would be helpful to conduct a genome-wide association study. However, because of the relatively small sample size with limited statistical power in this study, we performed a targeted association analysis for only one MetS-related SNP with MetS-related and heritable taxa. Second, our study cannot define the sequential causal relationship in this setting. Future studies using animal

models with gene knock-outs are warranted to confirm a causal link between *APOA5* genotype, alteration of gut microbiota, and MetS development.

In summary, we identified a specific set of gut microbial taxa that were significantly associated with MetS status and their relationship with host genetics. Our results suggest the possibility that an altered gut microbiota, which may be partially mediated by specific host genetic variants, can contribute to the development of MetS. It will be important to consider host genetics when developing future microbiome-targeted therapeutic strategies for metabolic disorders such as MetS.

CHAPTER IV.

HUMAN GENETICS AND SMOKING SHAPE THE AIRWAY MICROBIOTA

Introduction

Until recently, due to limitations in analytical methods, the airways of healthy humans were thought to be sterile (Marsland and Gollwitzer, 2014). However, recent studies using high-throughput next-generation sequencing (NGS) techniques have reported that the airways are not sterile (even in a healthy host) and harbor diverse microbial communities (Charlson *et al.*, 2011, Hilty *et al.*, 2010). Several investigations have shown that changes in the composition of the airway microbiota are associated with the development of chronic lung diseases such as asthma, chronic obstructive pulmonary disease (COPD), and cystic fibrosis (CF) (Erb-Downward *et al.*, 2011, Hilty *et al.*, 2010, van der Gast *et al.*, 2011). In addition, a decreased microbiota diversity has been reported in these diseases (Blainey *et al.*, 2012, Erb-Downward *et al.*, 2011). The airway microbiota plays an important role in host health and diseases, although the underlying mechanisms remain unclear. Therefore, it is important to characterize the "normal" airway microbiota in a large, healthy population to gain insights into the basic characteristics of airway microbiota, to identify causes of airway microbiota variations across healthy individuals, and to use the information as a reference for identifying specific microbial signatures of microbiota-related lung diseases.

Chronic lung diseases are caused by both genetic and environmental factors. Numerous epidemiological studies have reported that chronic lung diseases are influenced by various lifestyle factors. It is well-established that smoking is the main risk factor for the development of not only COPD including emphysema and chronic bronchitis but also asthma in adults (Løkke *et al.*, 2006, Lindberg *et al.*, 2005, Piipari *et al.*, 2004). For examples, a 25 year follow-up study showed that more than 25% of continuous smokers develop COPD (Løkke *et al.*, 2006). Lifestyle factors such as physical inactivity and alcohol consumption are also known to be associated with these diseases (Lieberoth *et al.*, 2012, Reilly *et al.*, 2008). In addition, several genome-wide

association studies have identified susceptibility loci that are associated with the diseases (Chen *et al.*, 2015, Cho *et al.*, 2014, Torgerson *et al.*, 2011).

Few studies to date have examined associations between host genetic or lifestyle factors and the airway microbiota. For example, it has been reported that the use of antibiotics, a class approach to control chronic respiratory infections in CF patients, is related to the decreased richness of the airway microbiota (Daniels *et al.*, 2013, Stokell *et al.*, 2015). Although smoking is another environmental factor that can lead to alterations in the composition of the airway microbiota by directly affecting the airway environment (Demling, 2008), at this time few studies have assessed the links between smoking and the airway microbiota (Charlson *et al.*, 2010, Erb-Downward *et al.*, 2011, Morris *et al.*, 2013, Segal *et al.*, 2013), and little is known regarding how smoking affects a normal balanced microbiota. One of these previous studies compared the bronchoalveolar lavage communities of healthy non-smokers (n=45) and smokers (n=19), but did not observe significant differences in lung microbiota (Morris *et al.*, 2013). These findings should be confirmed in a larger study. In terms of host genetic effects, host genetics play a role in shaping the gut microbial composition (Goodrich *et al.*, 2014). However, whether the airway microbiota of healthy individuals is influenced by host genetic factors has yet to be determined. In this context, our knowledge of the airway microbiota and its relationship with various host-associated factors remains limited.

In this study, to evaluate the effects of host genetics and lifestyle factors such as smoking, alcohol consumption, and physical activity on the airway microbiota composition, we comprehensively investigated associations of airway microbiota with host factors using a total of 257 sputum samples and their metadata from healthy Korean monozygotic (MZ) twins, dizygotic (DZ) twins, and their families.

Materials and Methods

Study subjects and sputum sample collection

The sputum samples and associated metadata were obtained from participants enrolled in the Healthy Twin Study in Korea (Sung *et al.*, 2006). Samples were excluded from this study if the participants had received antibiotic treatment or cold medication within the past three months. In addition, samples from participants with asthma or COPD were excluded. The final study subjects consisted of 257 individuals, including 74 pairs of MZ twins (n=148), 14 pairs of DZ twins (n=28), and their parents or siblings (n=81).

Prior to sputum induction, distilled water was used as a mouth rinse to minimize contamination of sputum samples with oral microbes. Serial expectorated sputum samples were collected in sterile plastic tubes and stored at -70°C until use. All subjects were asked to complete a questionnaire about smoking, alcohol consumption, and physical activity. The pack-year of smoking was calculated as packs per day multiplied by years of smoking. Weekly consumption of alcohol (g/week) was calculated as the frequency of drinking per week multiplied by the mean amount of alcohol consumption at each drinking session. Physical activity was assessed using the Korean version of the International Physical Activity Questionnaire (Booth *et al.*, 2003). Nicotine dependence was measured using the Fagerstrom Test of Nicotine Dependence (FTND) (Heatherton *et al.*, 1991). The study was approved by the Institutional Review Board of Samsung Medical Center (IRB file No. 2005-08-113), Busan Paik Hospital (IRB file No. 05-037), and Seoul National University (IRB file No. 144-2011-07-11). All participants provided written informed consent.

DNA extraction and 16S rRNA gene sequencing

The sputum samples were washed twice with cold phosphate-buffered saline (PBS), followed by DNA extraction using a PowerSoil DNA Isolation Kit (MO BIO

Laboratories, Carlsbad, CA, USA) according to the manufacturer's instructions. DNA was eluted in 50 μL of MoBio elution buffer and then stored at -20°C until use. The 16S rRNA genes were amplified using the 515F and 806R primers for the V4 region, as described previously (Caporaso *et al.*, 2012). PCR products were purified using MO BIO UltraClean PCR Clean-Up Kit (MO BIO Laboratories, Carlsbad, CA, USA), and subsequently quantified using the KAPA Library Quantification Kit (KAPA Biosystems, Woburn, MA, USA). Sequencing was performed on the MiSeq platform using a paired-end 2 × 300-bp reagent kit according to the manufacturer's instructions (Illumina, San Diego, CA, USA).

Sequence analysis

Sequences generated from the MiSeq run were processed using QIIME v1.8.0 (Caporaso *et al.*, 2010) and were clustered into 97% identity using a two-step open-reference operational taxonomic unit (OTU) picking method. Taxonomy was assigned to each OTU using the Ribosomal Database Project (RDP) classifier based on the 13_5 revision of the Greengenes database. Read counts from the OTU table were collapsed at six levels from domain to genus, and subsequently converted to relative abundances for further analysis. Unclassified taxa at a given taxonomic level were excluded from the analysis.

Bioinformatics analysis and statistical test

The Bray-Curtis distances between each sample were calculated using the genus-level abundance profiles to evaluate the similarity between microbial communities from MZ twin pairs, DZ twin pairs, family members of the twin pairs, and unrelated individuals belonging to the various twin families. The statistical significance of differences was evaluated using a two-sample t-test via 1,000 Monte Carlo permutations. For the test, 20% of the unrelated pairs selected randomly were used to reduce the number of unrelated pairs.

Heritability for each microbial taxon was estimated using a variance component method implemented in Sequential Oligogenic Linkage Analysis Routines (SOLAR) (Almasy and Blangero, 1998). Low-abundance taxa (average relative abundance < 0.1%) were excluded from this analysis. Heritability estimates (H2r) were calculated for inverse normalized microbial abundances, after controlling for age, sex, and pack-years of smoking. P-values were false discovery rate (FDR)-corrected for multiple testing.

Non-metric multidimensional scaling (NMDS) analysis was performed using MetaMDS function within the vegan package in R based on the Bray–Curtis distance measure obtained for the genus-level data. A total of 20 iterations were performed and the NMDS dimension with the lowest stress was retained for data visualization. Vectors of metadata were fitted onto the NMDS ordination plot using the envfit function of the vegan package.

Multivariate analysis was conducted using multivariate association by linear models (MaAsLin) (Morgan *et al.*, 2012) to test for associations of microbial abundances (at all taxonomic levels from domain to genus) with lifestyle factors, such as smoking (pack-years or FTND score), alcohol consumption, and physical activity. In this analysis, age and sex were used as fixed effects, and MZ twin ID and family ID were used as random effects. Only taxa with average abundances across all samples > 0.1% were included in the MaAsLin analysis. FDR-corrected p-values (q-values) of < 0.25 were considered significant. Univariate analysis was conducted to identify taxonomic features differentially abundant between never-smokers and ever-smokers using linear discriminant analysis effect size (LEfSe) (Segata *et al.*, 2011) under the condition $\alpha = 0.05$, with an LDA score of at least 2.

Alpha diversity indexes—the Chao1 index (richness) and the Shannon index (diversity)—were calculated with rarefied data (10,000 sequences per sample) using QIIME (Caporaso *et al.*, 2010). Wilcoxon rank sum test and Spearman's correlation test were used to evaluate the significance of the association between these index measures and metadata. The co-occurrence relationships between the relative abundances of taxa

(at the genus level) found to be significantly heritable or significantly associated with smoking were evaluated by calculating Spearman's ranked correlation. Networks of co-occurring taxa were visualized using the edge-weighted spring embedded layout algorithm in Cytoscape (Smoot *et al.*, 2011). In the network, each node represents a genus, and pairs of nodes were linked by edges (correlation coefficients) only when their FDR-corrected p-values were less than 0.25.

Results

Participant characteristics

A total of 257 sputum samples from individuals enrolled in the Healthy Korean Twin Study cohort was collected as described in a previous study (Gombojav *et al.*, 2013). This study population comprised 74 MZ twin pairs (n=148) and their parents or siblings (n=58), as well as 14 DZ twin pairs (n=28) and their parents or siblings (n=23). The average age of the subjects was 46.15 (±12.70) years, and 57.98% were female (Table 4.1). All subjects had not taken antibiotics within three months prior to sampling and had no current symptoms of chronic lung diseases, such as asthma or COPD. Demographic and lifestyle characteristics of the study participants are shown in Table 4.1.

Table 4.1 Characteristics of participants in this study.

Variable	Overall (n=257)	MZ twin		DZ twin	
		Twin pairs (n=148)	Family members (n=58)	Twin pairs (n=28)	Family members (n=23)
Females, n (%)	149 (57.98)	84 (56.76)	34 (58.62)	16 (57.14)	15 (65.22)
Age (years)	46.15 (12.70)	41.22 (8.67)	56.93 (14.27)	40.5 (7.31)	58.9 (11.58)
BMI (kg/m^2)	23.63 (3.21)	23.33 (2.86)	23.95 (3.59)	23.19 (3.55)	24.7 (3.53)
Smoking (pack-years)[§]	12.25 (12.78)	11.46 (12.42)	14.61 (15.49)	10.65 (8.24)	5 (10.03)
Alcohol consumption (g/week)[#]	121.04 (257.00)	116.26 (211.93)	153.15 (442.45)	94.37 (111.37)	136.79 (156.5)
Physical activity (MET-min/week)[$]	5885.75 (9827.53)	5233.04 (9854.18)	6131.18 (8760.54)	6969.73 (12388.92)	8865.05 (12055.24)

Data are presented as means (standard deviation) unless otherwise stated.

Missing data (numbers for overall, MZ twin pairs, family members of MZ twins, DZ twin pairs, and family members of DZ twins, respectively)

[§]One hundred and fifty-two subjects with missing data (152, 84, 32, 20, 16);

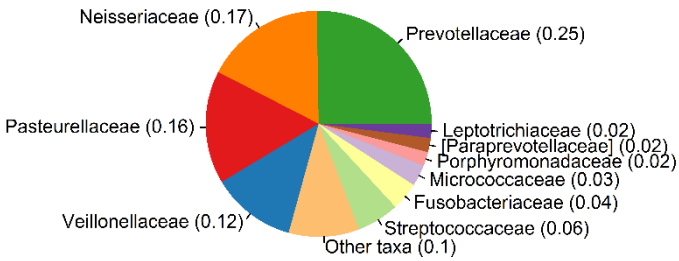
[#]Sixty-four subjects with missing data (64, 27, 23, 4, 10);

[$]Thirty-seven subjects with missing data (37, 23, 7, 4, 3).

Core airway microbiota of a healthy population

A mean of 46,311 sequences (± 30,529 SD) per sample were obtained by 16S rRNA gene sequencing from sputum samples collected from each subject. Proteobacteria (34%), Bacteroidetes (32%), Firmicutes (21%), Fusobacteria (6%), and Actinobacteria (4%) constituted the five most abundant bacterial phyla. At the family level, Prevotellaceae (25%), Neisseriaceae (17%), Pasteurellaceae (16%), and Veillonellaceae (12%) were the most abundant families (Figure 4.1A). These families were present in all samples, and constituted the family-level core airway microbiota. At the genus level, *Prevotella* (25%) and *Neisseria* (16%), followed by *Haemophilus* (11%), *Veillonella* (11%), and *Streptococcus* (6%) were most abundant in this healthy population, and were present as members of the genus-level core microbiota (Figure 4.1B). Notably, *Prevotella* showed substantial inter-individual variation. All members of the core airway microbiota (> 1% average relative abundance and 100% prevalence) at the genus and family levels are represented in Figure 4.1.

(A)



(B)

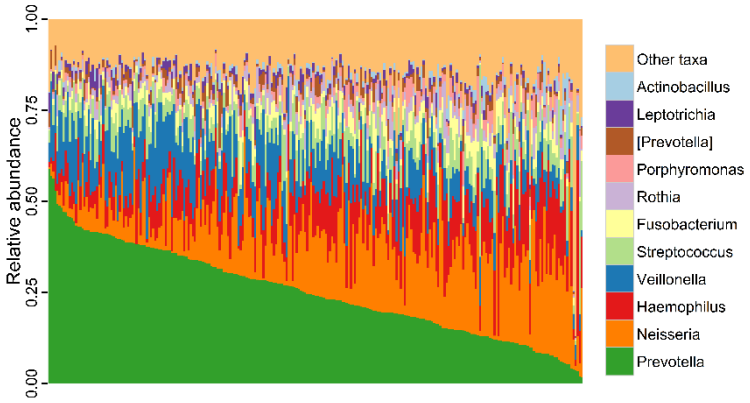


Figure 4.1 Composition of the airway microbiota. (A) Pie chart of the average relative abundances of the families (> 1% average relative abundance and 100% prevalence). (B) Bar chart of relative abundances of the genera across each sampled airway microbiome (> 1% average relative abundance and 100% prevalence).

Host genetic effects on the airway microbiota

We investigated whether host genetics influence airway microbiota. By comparing Bray-Curtis distances between MZ twin pairs, DZ twin pairs, family members of the twin pairs, and unrelated individuals belonging to the various twin families, we found that the airway microbiota of MZ twin pairs were significantly more similar to one another than were those of DZ twin pairs (p-value=0.046, two sample t-test via 1,000 Monte Carlo permutations; Figure 4.2A). As expected, MZ twin pairs had more-similar microbial community structures than family members and unrelated individuals (p-value=0.004 and p-value< 2.2e-16; Figure 4.2A). This result suggested that the overall community structure of airway microbiota could be partially modulated by host genetics.

To identify specific microbial taxa with heritability, we estimated the heritability of each taxa based on a variance component model using SOLAR (Almasy and Blangero, 1998). Among the 71 tested taxa with > 0.1 % average relative abundance, 36 taxa (50.7%) were significantly heritable at a q-value of < 0.25. The heritability estimates of these taxa ranged from 0.17 to 0.70 (Appendix C). At the genus level, we observed significant moderate (H2r = 0.3~0.6) to high (H2r > 0.6) heritability for *Providencia* (H2r ± SE; 0.70±0.06), *Bacteroides* (0.55±0.07), *Prevotella* (0.48±0.40), *Haemophilus* (0.43±0.15), *Gemella* (0.42±0.09), *Porphyromonas* (0.36±0.09), and *Rothia* (0.36±0.15) (Figure 4.2B). Collectively, these data demonstrate that host genetic factors could specifically affect a portion of the airway microbiota.

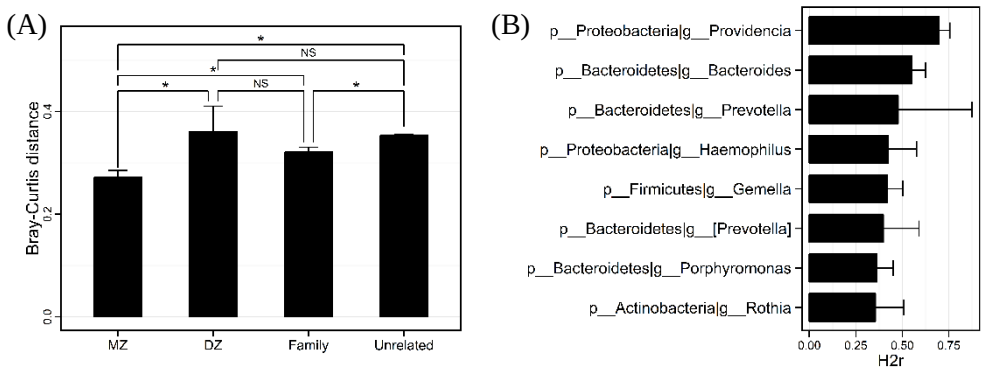


Figure 4.2 Influences of host genetic factors on the airway microbiota. (A) Comparison of the airway microbiota from monozygotic (MZ) twin pairs, dizygotic (DZ) twin pairs, family members of the twin pairs, and unrelated individuals belonging to the various twin families. Bray-Curtis distances were calculated with genus-level relative abundance profiles (mean ± s.e.m.; *P < 0.05 for two sample t-test via 1,000 Monte Carlo permutations). (B) Heritability estimates (H2r) of airway microbial taxa (mean ± s.e.m.). Genera with significant heritability (H2r > 0.3 and FDR-corrected p-values < 0.25) are shown.

Relationship between smoking and airway microbiota

We first investigated the relationship between the overall airway microbial community structure and lifestyle factors (smoking history (pack-years), alcohol drinking (g/week), and physical activity (MET-min/week)) by conducting NMDS analysis. We found that pack-years of smoking was significantly correlated with the NMDS ordination of the microbial community structure (R^2=0.110, p-value=0.027); increasing pack-year values was represented by a vector pointing to the right of the NMDS plot (Figure 4.3A). The remaining lifestyle factors were not significantly correlated with the NMDS ordination of the microbial community structure (alcohol consumption: R^2=0.086, p-value=0.054; physical activity: R^2=0.046, p-value=0.195). Additionally, the FTND score, a commonly used measure of nicotine dependence, was significantly correlated with the NMDS ordination (R^2=0.161, p-value=0.003) (Figure 4.3A). We categorized the individuals into ever-smokers and never-smokers based on the pack-year value: Individuals with a > 0 pack-year smoking history were defined as ever-smokers, whereas those with pack-year values of zero were defined as never-smokers. This categorical variable also showed a significant correlation with the NMDS ordination (R^2=0.073, p-value=0.012) (Figure 4.3A). Therefore, our data showed that smoking significantly influenced the structure of the airway microbial communities.

To identify specific microbial taxa significantly associated with lifestyle factors while accounting for age, sex, and family structures, we performed the MaAsLin analysis (Morgan *et al.*, 2012). Among the lifestyle factors, the variable of pack-year smoking history showed the highest number of significant associations with microbial taxa (n=14): pack-year of smoking was positively associated with taxa from the Firmicutes phylum, the Clostridiales order, the Veillonellaceae family, and the *Veillonella* genus, while this variable was inversely associated with taxa from the Gammaproteobacteria class, including the *Actinobacillus* and *Haemophilus* genera (Figure 4.3B and Table 4.2). In addition, we observed significant associations of specific microbial taxa with alcohol consumption (n=5) and physical activity (n=1). For example, alcohol consumption was

positively associated with taxa from Actinobacteria *Rothia*, and physical activity was negatively associated with the Firmicutes Lachnospiraceae, respectively (Table 4.2). When testing for an association of the microbial taxa with the FTND, instead of a pack-year variable, we found that higher FTND scores were associated with increased abundances of taxa from the Veillonellaceae family, including the *Veillonella* and *Megasphaera* genera (Figure 4.3B and Table 4.3). The *Haemophilus* and taxa to which this genus belongs were negatively associated with FTND score (Table 4.3).

To confirm the results of multivariate analyses, we performed univariate analysis using LEfSe (Segata *et al.*, 2011) with respect to smoking history (ever- vs. never-smokers). A total of 24 microbial taxa were significantly differentially abundant between never-smokers and ever-smokers (Figure 4.4 and Table 4.4). The genera including *Veillonella*, *Megasphaera*, and *Prevotella* were significantly enriched in ever-smokers, while *Haemophilus* and *Actinobacillus* were increased in never-smokers. This observation was consistent with multivariate analysis. These results suggested that smoking is a major environmental factor that not only drives the difference in microbial community structure between samples but also affects the abundance of specific microbial taxa.

We additionally explored the association between lifestyle variables and alpha diversity indexes (Chao1 richness and Shannon diversity index). There were no significant correlations between the lifestyle factors and alpha diversity (Spearman's correlation; $\rho = 0.03 \sim 0.07$, p-value > 0.05, Figure 4.5). In agreement with this, both richness and diversity estimates were not significantly different regarding age or sex (Spearman's correlation; $\rho = 0.06$, p-value > 0.05, Wilcoxon rank sum test; p-value > 0.05, Figure 4.5).

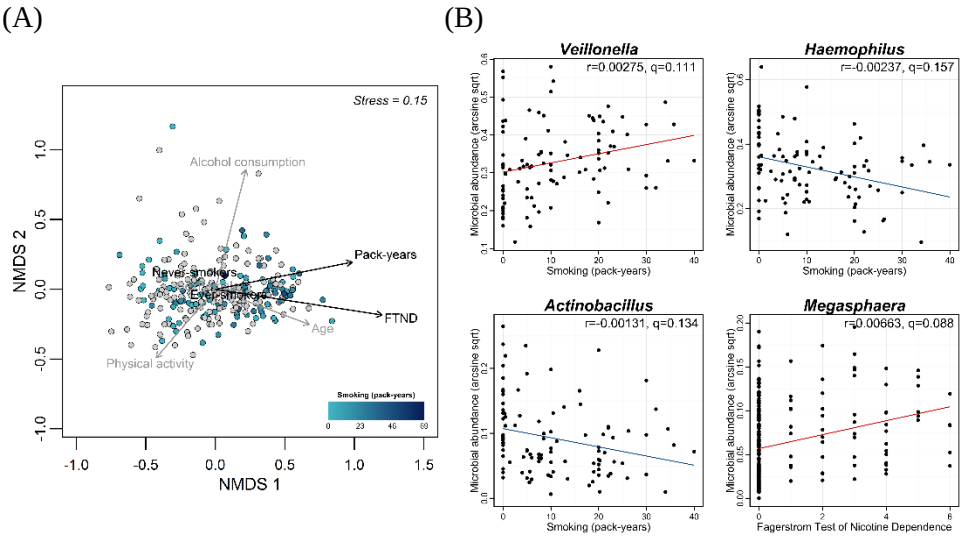


Figure 4.3 Associations of lifestyle factors with airway microbiota. (A) Non-metric multidimensional scaling (NMDS) plot of genus composition. Points represent samples that are colored according to pack-years of smoking (a gradient from light blue to dark blue). Gray points indicate samples with missing pack-year values. Vectors represent correlations between community composition and metadata. The black arrows indicate significant correlations with the ordination, whereas the gray arrows indicate non-significant correlations. (B) Significant associations between smoking and microbial taxa. The r-coefficient and q-value shown in each plot were determined by MaAsLin analysis.

Table 4.2 Complete list of significant associations of lifestyle factors with microbial abundances.

Variable	Feature	Coefficient	P.value	Q.value
Alcohol consumption (g/week)	p__Actinobacteria	0.000115655	0.03146	0.21478
Alcohol consumption (g/week)	p__Actinobacteria|c__Actinobacteria	0.000117847	0.02965	0.21052
Alcohol consumption (g/week)	p__Actinobacteria|o__Actinomycetales	0.000117936	0.02919	0.21052
Alcohol consumption (g/week)	p__Actinobacteria|f__Micrococcaceae	0.000124183	0.01729	0.17048
Alcohol consumption (g/week)	p__Actinobacteria|g__Rothia	0.00012418	0.01729	0.17048
Smoking (Pack-years)	p__Firmicutes	0.002753123	0.00433	0.11086
Smoking (Pack-years)	p__Firmicutes|c__Clostridia	0.002686105	0.01056	0.12108
Smoking (Pack-years)	p__Firmicutes|o__Clostridiales	0.002685454	0.01057	0.12108
Physical activity (MET-min/week)	p__Firmicutes|f__Lachnospiraceae	-1.05E-06	0.02062	0.18653
Smoking (Pack-years)	p__Firmicutes|f__Peptostreptococcaceae	-0.0002458	0.02276	0.19239
Smoking (Pack-years)	p__Firmicutes|f__Veillonellaceae	0.002872027	0.00715	0.11086
Smoking (Pack-years)	p__Firmicutes|g__Veillonella	0.002752688	0.00548	0.11086
Smoking (Pack-years)	p__Fusobacteria|o__Fusobacteriales	-0.00159946	0.02907	0.21052
Smoking (Pack-years)	p__Fusobacteria|f__Fusobacteriaceae	-0.00142758	0.03289	0.21624
Smoking (Pack-years)	p__Fusobacteria|g__Fusobacterium	-0.00142538	0.03251	0.21624
Smoking (Pack-years)	p__Proteobacteria	-0.00375528	0.02155	0.1866
Smoking (Pack-years)	p__Proteobacteria|c__Gammaproteobacteria	-0.00328716	0.00363	0.11086
Smoking (Pack-years)	p__Proteobacteria|f__Pasteurellaceae	-0.00290249	0.00818	0.11086
Smoking (Pack-years)	p__Proteobacteria|g__Actinobacillus	-0.00130816	0.01212	0.13444
Smoking (Pack-years)	p__Proteobacteria|g__Haemophilus	-0.00236609	0.01506	0.15724

Table 4.3 Significant associations of Fagerstrom Test of Nicotine Dependence (FTND) scores with microbial abundances.

Variable	Feature	Coefficient	P.value	Q.value
Smoking (FTND)	p__Firmicutes|g__Megasphaera	0.007	0.00148	0.08767
Smoking (FTND)	p__Proteobacteria|c__Gammaproteobacteria	-0.014	0.00911	0.17201
Smoking (FTND)	p__Proteobacteria|g__Haemophilus	-0.012	0.01026	0.17201
Smoking (FTND)	p__Firmicutes|f__Veillonellaceae	0.013	0.01172	0.17201
Smoking (FTND)	p__Proteobacteria|f__Pasteurellaceae	-0.013	0.01269	0.17201
Smoking (FTND)	p__Firmicutes|c__Clostridia	0.012	0.01464	0.17201
Smoking (FTND)	p__Firmicutes|o__Clostridiales	0.012	0.01465	0.17201
Smoking (FTND)	p__Firmicutes|g__Veillonella	0.011	0.02511	0.21743
Smoking (FTND)	p__Proteobacteria	-0.017	0.02845	0.23488

FTND values were considered a continuous variable.

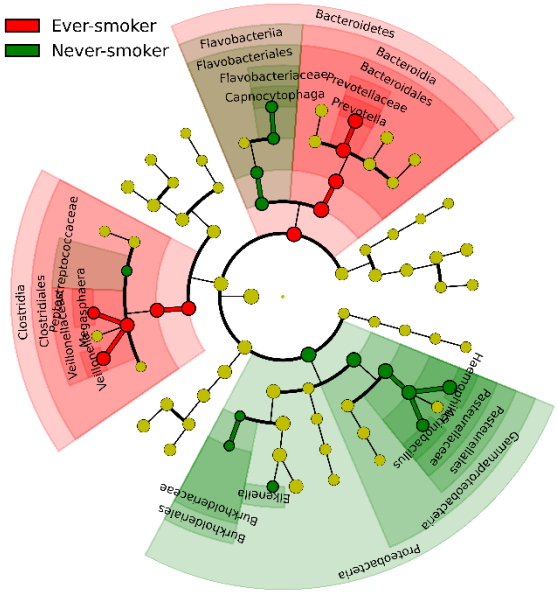


Figure 4.4 Significant differences in the composition of the airway microbiota between ever- and never-smokers. Circular cladogram obtained from univariate analysis (LDA Effect Size) showing the taxa enriched in the airway microbiota of ever- (red) and never- (green) smokers. Each ring from the center to the periphery represents phylogenetic levels from phylum to genus.

Table 4.4 Significant differences in the relative abundances of the airway microbiota between ever- and never-smokers.

Taxon	Ever-smokers		Never-smokers		Fold change	LDA score (log 10)
	Mean	S.E.M	Mean	S.E.M		
p__Bacteroidetes	0.3504	0.0126	0.2729	0.0217	1.2839	4.5764
p__Bacteroidetes|o__Bacteroidales	0.3427	0.0129	0.2607	0.0224	1.3146	4.5989
p__Bacteroidetes|f__Prevotellaceae	0.2912	0.0132	0.2068	0.0239	1.4084	4.6210
p__Bacteroidetes|g__Prevotella	0.2912	0.0132	0.2067	0.0239	1.4085	4.6209
p__Firmicutes|c__Clostridia	0.1439	0.0075	0.1185	0.0183	1.2146	4.2333
p__Firmicutes|o__Clostridiales	0.1438	0.0075	0.1184	0.0183	1.2145	4.2327
p__Firmicutes|f__Veillonellaceae	0.1342	0.0074	0.1084	0.0182	1.2384	4.1400
p__Firmicutes|g__Veillonella	0.1171	0.0064	0.0976	0.0159	1.2001	4.0399
p__Firmicutes|g__Megasphaera	0.0085	0.0009	0.0045	0.0015	1.8822	3.2967
p__Proteobacteria	0.2976	0.0164	0.4041	0.0317	0.7366	4.6943
p__Proteobacteria|c__Gammaproteobacteria	0.1405	0.0087	0.2236	0.0195	0.6283	4.5739
p__Proteobacteria|f__Pasteurellaceae	0.1372	0.0087	0.1982	0.0171	0.6922	4.4739
p__Proteobacteria|g__Haemophilus	0.0982	0.0064	0.1472	0.0140	0.6674	4.3633
p__Proteobacteria|g__Actinobacillus	0.0126	0.0032	0.0179	0.0035	0.7044	3.5777
p__Bacteroidetes|o__Flavobacteriales	0.0071	0.0009	0.0117	0.0020	0.6112	4.0445
p__Bacteroidetes|f__Flavobacteriaceae	0.0054	0.0008	0.0092	0.0019	0.5860	3.8296
p__Bacteroidetes|g__Capnocytophaga	0.0054	0.0008	0.0092	0.0019	0.5856	3.7711
p__Proteobacteria|g__Eikenella	0.0013	0.0001	0.0051	0.0014	0.2492	3.3247
p__Firmicutes|f__Peptostreptococcaceae	0.0008	0.0001	0.0012	0.0003	0.6628	2.8452
p__Proteobacteria|o__Burkholderiales	0.0008	0.0001	0.0011	0.0001	0.7931	2.6930
p__Proteobacteria|f__Burkholderiaceae	0.0007	0.0001	0.0010	0.0001	0.7153	2.7543

LDA scores (log 10) were obtained from univariate analysis (LDA Effect Size).

Fold change for each taxon was calculated by dividing the mean abundance in the ever-smokers by that of the never-smokers.

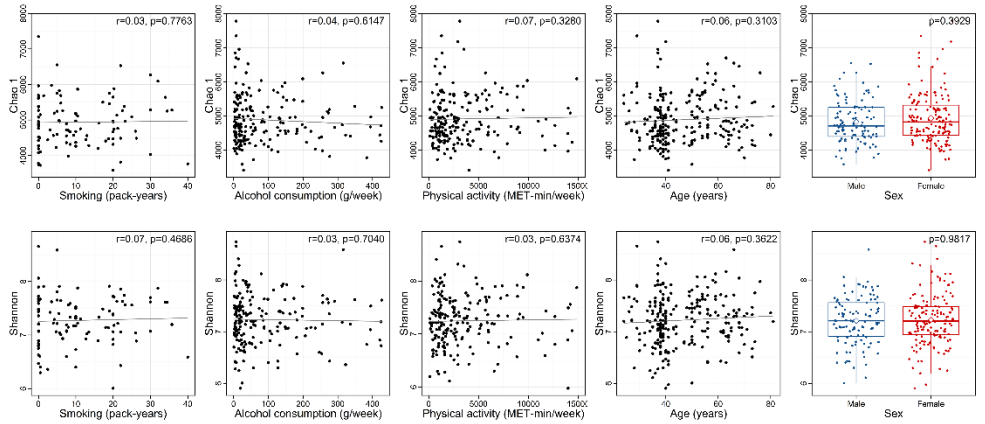


Figure 4.5 Choa1 richness (upper panel) and Shannon diversity index (lower panel) according to lifestyle factors (smoking, alcohol consumption, and physical activity), age, and sex. The alpha diversity indexes were calculated from rarefied OTU table. The r-coefficient and p-value shown in each scatter plot were calculated by Spearman's rank correlation. The p value shown in each box plot was calculated by Wilcoxon rank-sum test.

The influence of smoking and host genetics on specific taxa of the airway microbiota

Our findings indicate that both host genetics and smoking influence specific taxa of the airway microbiota. To increase our understanding of the interactions among airway microbial communities, smoking, and host genetics, we performed a further analysis by combining the results of the following two analyses: SOLAR analysis (estimating the heritability of specific taxa; Appendix C) and MaAsLin analysis (testing for the association between smoking and specific taxa; Table 4.2). For the genera shown to be significantly associated with pack-year of smoking (hereafter, "smoking-associated genera") or the genera with significant moderate to high heritability (hereafter, "host genetics-associated genera"), the q-values from MaAsLin (x-axis) were plotted as individual points against the corresponding q-values from SOLAR (y-axis) in Figure 4.6A. In most cases, the genera showed significant associations with either smoking or host genetics. *Haemophilus* was the only genus found to be significantly associated with both factors.

To reveal relationships among the smoking-associated genera and the host genetic-associated genera, we performed co-occurrence analysis based on Spearman's rank correlations. In this analysis, we also included genera that were significantly associated with the FTND score or the categorical variable of smoking history (ever- vs. never-smokers). There were 65 significant correlations that had a q-value of < 0.25 (Figure 4.6B). We observed that the smoking-associated genera co-occurred according to the tendency toward smoking. For example, *Veillonella*, which was positively associated with pack-year values, had positive correlations with *Megasphaera* and *Prevotella* (Spearman's correlation; $\rho = 0.64$ and 0.54, respectively). Other smoking-associated genera, such as *Haemophilus* and *Actinobacillus*, which were negatively associated with pack-year values, co-occurred together (Spearman's correlation; $\rho = 0.31$). On the other hand, *Veillonella* was anti-correlated with *Haemophilus* and *Actinobacillus* (Spearman's correlation; $\rho = -0.23$ and -0.28, respectively). Additionally, we observed that the genera

associated only with host genetics were closely linked, forming a *Rothia*-centered cluster. A strong positive correlation was observed between the two most-heritable genera, *Providencia* and *Bacteroides* (Spearman's correlation; $\rho = 0.66$), and these genera were positively correlated with *Rothia* (Figure 4.6B). The data demonstrated that relationships among members of the airway microbiota are affected by both smoking and host genetics.

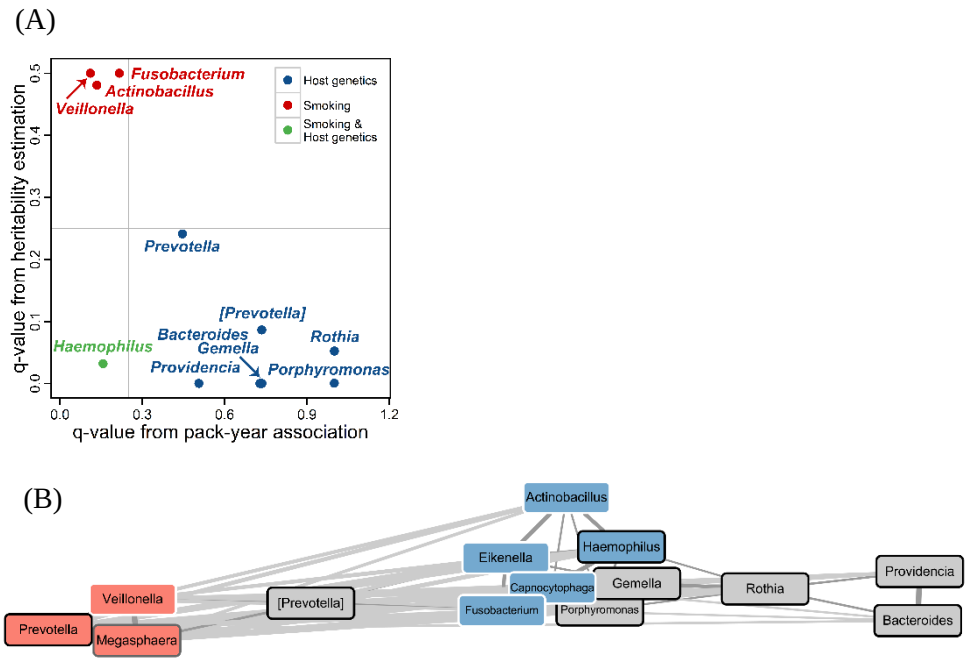


Figure 4.6 Effects of smoking and host genetics on specific microbial taxa. (A) Scatter plot of FDR-corrected p-values (q-values) for the association of each genus with pack-years of smoking (x-axis; obtained from MaAsLin analysis) versus q-values for the heritability of each genus (y-axis; obtained from SOLAR analysis). Red, blue, and green points represent genera that are significantly associated with pack-years of smoking, host genetics, and both, respectively. Here, q-values < 0.25 were considered significant. (B) Co-occurrence and co-exclusion relationships among the genera shown to be significantly associated with smoking (pack-years, FTND, or ever- vs. never-smokers) or to have significant heritability (H2r > 0.3). Nodes represent genera. Nodes are colored according to the associations between genus and smoking: Pink and light blue nodes indicate genera positively and negatively associated with smoking, respectively. Gray nodes indicate genera not significantly associated with smoking. The border of each node is colored according to the significance of the heritability (black: q-value < 0.25 and H2r > 0.3; dark gray: q-value < 0.25 and H2r < 0.3; white: q-value ≥ 0.25). Edges

represent significant positive (dark gray) or negative (light gray) correlations between the nodes they connect. The width of the edge represents the degree of correlations.

Discussion

In this study, we showed that both host genetics and smoking influence the abundance of specific taxa of the airway microbiota in healthy adult populations. This finding is important because specific taxa of the airway microbiota and associated host factors must be identified to understand the underlying mechanism of, and develop methods to control, dysbiosis of the airway microbiota, which may have an important role in the development of lung diseases.

Most previous studies on the airway microbiota have focused on identifying differences in the microbiota composition of patients with lung diseases and healthy controls. By contrast, in this study, we explored the characteristics of a normal airway microbiota in healthy populations. Our relatively large sample size (n=257) allowed us to determine the core airway microbiota and to investigate associations of various lifestyle factors with the airway microbiota with high precision. Furthermore, our study population, consisting of twins and their families, allowed us to estimate the heritability of each member of the microbiota. Therefore, we detected independent effects of lifestyle factors on the airway microbiota taking into consideration host genetic factors in this study.

We characterized the composition of the normal airway microbiota and identified the core members, which included *Prevotella*, *Neisseria*, *Veillonella*, and *Streptococcus*, which were present in all samples at > 1% average relative abundance (Figure 4.1). Our data are consistent with previous studies using sputum samples from a small number of healthy subjects (Cui *et al.*, 2012, Wu *et al.*, 2013). Interestingly, most of the core airway microbiota in our samples corresponded to predominant taxa of the oral microbiota (Nasidze *et al.*, 2009, Zaura *et al.*, 2009). A previous study reported differences in the T-RFLP profiles between sputum and oral mouthwash samples from CF subjects (Rogers *et al.*, 2006). In our study, we washed the mouth of the subjects prior to

obtaining sputum samples. Therefore, while contamination with oral microbes was possible, the degree of any such contamination should be minimal. Because sputum samples can be obtained less invasively and more readily, sputum would be the most suitable sample for analyzing the airway microbiota of a large, healthy population.

Host genetics was an important factor influencing the composition of the airway microbiota. The overall microbiota structures from MZ twin pairs were significantly more similar compared to those from DZ twin pairs, family members, or unrelated individuals. These results suggest the airway microbiota to be at least in part genetically influenced and have high heritability (Figure 4.2A). As shown in Figure 4.2B, many specific airway microbial taxa showed significant heritability estimates. When we estimated the heritability of specific microbes, we adjusted for age, sex, and smoking history. Therefore, the high heritability of some members of the airway microbiota remains valid regardless of these factors. Among them, *Providencia* (belonging to Proteobacteria) was the most highly heritable taxon (Figure 4.2B). Five species of *Providencia* have been isolated from various human specimens, such as urine, sputum, and stool (O'Hara *et al.*, 2000). Although *P. stuartii* has long been recognized as an opportunistic pathogen responsible for urinary tract infections in long-term care patients with chronic indwelling urinary catheters (Nicolle, 2002), the effect of *Providencia* on the human airway remains largely unknown. Therefore, how *Providencia* is influenced by host genetics should be further investigated. Among the other airway microbiota taxa, *Bacteroides* and *Rothia* exhibited significant heritability. This finding is interesting because these taxa were also identified in the gut and were found to be highly heritable in UK twins (Goodrich *et al.*, 2014). Therefore, these two taxa may be specifically influenced by host genetic factors regardless of the body site in which they reside. There may be common phenotypic factors between bronchial and intestinal epithelial cells. Further studies are required to determine whether the heritability of these taxa can also be observed at other body sites and to identify the host alleles that contribute to the heritability. Goodrich *et al.* reported that the Christensenellaceae family had the highest

heritability among gut microbial taxa (Goodrich *et al.*, 2014). However, we did not estimate the heritability of this family because it was present at a very low prevalence (0.05) and low average relative abundance (2E-06) in the airway microbiota. Hampton *et al.* reported that the airway microbiota of 13 pediatric CF patients was susceptible to environmental influences rather than host genetic effects (Hampton *et al.*, 2014). This study suggests that both lifestyle and disease status can affect the composition of the airway microbiota.

Among the lifestyle factors investigated, smoking had the strongest effect on the overall microbial community structure (Figure 4.3A). This result is reasonable because cigarette smoke contains a wide range of components—including carcinogens, toxins, and oxidants (Smith and Hansch, 2000)—which can alter the airway environment by inducing cellular damage and inflammation (Stämpfli and Anderson, 2009). We further identified a specific set of microbial taxa significantly associated with smoking. Interestingly, we observed that the pack-year value was positively associated with abundance of the *Veillonella* genus and all taxa to which this genus belongs (Firmicutes; Clostridia; Clostridiales; Veillonellaceae) (Table 4.2). Changes in the abundances of the *Veillonella* genus were similar to those reported in the study by Charlson and colleagues (Charlson *et al.*, 2010) that compared the oropharynx of 29 cigarette smokers and 33 non-smokers, and were confirmed by both MaAsLin multivariate analysis using the FTND score and LEfSe univariate analysis using a categorical variable (ever- vs. never-smokers). The *Veillonella* genus is part of the normal flora of the oral, respiratory, and intestinal tracts of humans. A positive correlation of the relative abundance of *Veillonella* in the lung microbiota with exhaled nitric oxide (a non-invasive measure of airway inflammation) in asymptomatic individuals has been reported (Segal *et al.*, 2013). *Veillonella* is a Gram-negative bacterium that can induce an inflammatory host response by producing lipopolysaccharides (LPS) (Delwiche *et al.*, 1985). We also found that *Actinobacillus* and *Haemophilus* were present at greater abundance in individuals with lower pack-year values (Figure 4.3B and Table 4.2). This result is consistent with a

previous report (Charlson *et al.*, 2010). Interestingly, non-typeable *H. influenza* is a well-known pathogen that causes respiratory tract infections in COPD patients (Sethi, 2010), but the abundance of *Haemophilus* decreased with increasing pack-year values in the present study. Since members of the *Haemophilus* genus are common resident bacteria of the human airways (Fink and Geme, 2006), further studies are required to explore the effect of smoking on specific species or strains of this genus. These observations suggest that smoking alters the abundance of specific members of the airway microbiota, such as *Veillonella*, which can lead to an inflammatory condition or an environment in which pathogenic bacteria can thrive, and subsequently contribute to the development of smoking-related lung diseases.

In the correlation network, we observed a significant negative relationship between the two smoking-associated taxa, *Haemophilus* (Proteobacteria) and *Veillonella* (Firmicutes) (Figure 4.6B). Since we used Spearman's correlation measures for the relative abundances of the taxa, a high abundance of one member of the microbiota can be accompanied by a reduction in the relative abundances of other members. However, this co-exclusion relationship suggests that these taxa require different environmental conditions to survive. Indeed, *Haemophilus* is a Gram-negative, aerobic/facultative anaerobic bacterium belonging to the Pasteurellaceae family (Foster *et al.*, 2011), whereas *Veillonella* is a genus of anaerobic Gram-negative cocci (Rogosa, 1964). Therefore, changes in oxygen levels of the airway environment caused directly by smoking or indirectly by smoking-related host responses may play a role in their co-exclusion relationships, although details on the mechanisms are not yet available. In addition, the most heritable genus, *Providencia*, was present together with the second-most heritable genus, *Bacteroides*, in the co-occurrence network (Figure 4.6B). These taxa were present at low relative abundances (0.11–0.47%) in the airway microbiota, and each belongs to one of two distinct phyla. Thus, our result suggests that these minor, heritable taxa may interact in a cooperative manner, and have a preference for a similar airway environment that is under the influence of host genetics. By network analysis,

we showed that the relationships among microbial taxa are closely associated with not only smoking but also host genetics.

In summary, the data presented here indicate that both host genetics and lifestyle contribute to determining the composition of the normal airway microbiota of healthy adults. Cigarette smoking was found the lifestyle factor with the greatest influence on the composition of the airway microbiota. Further studies using longitudinal sample sets are required to understand the causative associations of smoking with specific taxa and to identify the health consequences of smoking-related changes in the airway microbiota. Our findings linking host genetics, lifestyle, and their airway microbiota can be used to establish future therapeutic biological markers for microbiome-mediated lung diseases.

CHAPTER V.

CONCLUSIONS

Summary and Conclusions

Under normal circumstances, the human microbiome is involved in a wide range of host-related processes including nutrient metabolism, development and maintenance of the immune system, and protection against pathogens. The host's intrinsic factors (such as age, sex, and host genetics) or extrinsic environmental factors (such as antibiotic treatment, diet, and lifestyle) can continually influence the composition and function of the human microbiota. In this thesis, three studies were performed to understand interactions between the human microbiome, the host's intrinsic and extrinsic factors, and health and diseases.

First, we characterized the enterotypes of healthy Koreans. We identified that 36 gut microbiota of healthy Korean twins can be categorized into two enterotypes based on the composition of their gut microbial communities. Our longitudinal samples of monozygotic (MZ) twins showed that the enterotype of each individual was generally stable over time and that the enterotype might be influenced by the host genetics to some extent. In addition, our data suggested that host properties such as long-term dietary patterns and a particular clinical biomarker could be important determinants of an individual's enterotype.

Second, we investigated the host genetic influences on the gut microbiota in metabolic syndrome (MetS). We characterized the gut microbiota of 655 MZ twins, dizygotic twins, and their families, of which approximately 18% had MetS. We found that there were significant differences in the composition of the gut microbiota between healthy and MetS groups. Among the gut microbes associated with MetS status, *Bifidobacterium* exhibited high heritability and its reduced abundance was significantly linked to the minor allele at the apolipoprotein A-V SNP rs651821, which is associated with triglyceride level and MetS. This result indicates that altered gut microbiota mediated by a specific host genotype can contribute to the development of MetS.

Third, we assessed the influences of lifestyle and host genetic factors on the normal airway microbiota. We identified the core airway microbiota from 257 sputum samples of healthy Korean twins and their families. Among the lifestyle factors, cigarette smoking (pack-year of smoking) had the strongest effect on the airway microbiota. Several taxa including *Providencia* and *Bacteroides* were found to be highly heritable. The network analysis showed that the composition of the airway microbiota is shaped by complex interactions among cigarette smoking, host genetics, and the relationships between community members.

These studies have several significance. The Korean population has different host genetics, dietary habits, and cultural background from the Western populations. Because most of large-scale human microbiome studies have been conducted in populations from the United States and Europe, we needed to generate the human microbiome data in a Korean population. Using a total of 948 fecal and sputum samples from Korean twins and their families, we intensively investigated the gut and airway microbiota. By comparing the gut microbiota between healthy and MetS individuals, we suggested the gut microbial taxa that can be used as diagnostic indicator or therapeutic target for MetS. In addition, to our knowledge, this study is the first to investigate the effect of human SNPs on specific gut microbiota, which is strongly associated with MetS status. Based on twin-family design, we were able to broaden our understanding of host genetic effects on the gut and airway microbiota by estimating the heritability of microbial taxa.

Overall, these data suggest that the human microbial composition is specifically modulated by both host genetics and extrinsic factors and that the altered microbiota mediated by these factors may be associated with increased risk of microbiome-related diseases such as MetS and airway inflammation. Therefore, our results emphasize the importance of personalized approaches in the prevention and treatment of these diseases. Future studies using animal models or longitudinal sample sets are needed to unveil causal relationships between host-related factors, specific microbes, and host phenotypes we observed here.

Bibliography

Abubucker S, Segata N, Goll J, Schubert AM, Izard J, Cantarel BL *et al.* (2012). Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Computational Biology* **8:** e1002358.

Ahn J, Sinha R, Pei Z, Dominianni C, Wu J, Shi J *et al.* (2013). Human gut microbiome and risk for colorectal cancer. *Journal of the National Cancer Institute* **105:** 1907-1911.

Ahn Y, Kwon E, Shim J, Park M, Joo Y, Kimm K *et al.* (2007). Validation and reproducibility of food frequency questionnaire for Korean genome epidemiologic study. *European Journal of Clinical Nutrition* **61:** 1435-1441.

Alexandra D, Christophe L, Van TP, Michael BZ, Christophe Dh, Annick B-D *et al.* (2014). Iron supplementation promotes gut microbiota metabolic activity but not colitis markers in human gut microbiota-associated rats. *British Journal of Nutrition* **111:** 2135-2145.

Almasy L and Blangero J (1998). Multipoint quantitative-trait linkage analysis in general pedigrees. *The American Journal of Human Genetics* **62:** 1198-1211.

Anderson MJ (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology* **26:** 32-46.

Angelakis E, Armougom F, Million M, Raoult D (2012). The relationship between gut microbiota and weight gain in humans. *Future Microbiology* **7:** 91-109.

Armougom F, Henry M, Vialettes B, Raccah D, Raoult D (2009). Monitoring bacterial community of human gut microbiota reveals an increase in Lactobacillus in obese patients and Methanogens in anorexic patients. *PLoS ONE* **4:** e7125.

Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende D *et al.* (2011). Enterotypes of the human gut microbiome. *Nature* **473:** 174-180.

Arumugam M, Raes J, Pelletier E, Paslier D, Yamada T, Mende DR *et al.* (2014). Addendum: Enterotypes of the human gut microbiome. *Nature* **506:** 516-516.

Ateequr R, Christian S, Olga G, Robert H, Stephan O, John FB *et al.* (2011). Nod2 is essential for temporal development of intestinal microbial communities. *Gut* **60:** 1354-1362.

Bäckhed F, Ding H, Wang T, Hooper LV, Koh GY, Nagy A *et al.* (2004). The gut microbiota as an environmental factor that regulates fat storage. *Proceedings of the National Academy of Sciences* **101:** 15718-15723.

Balamurugana R, Georgea G, Kabeerdossa J, Hepsibaa J, Chandragunasekarana A, Ramakrishnaa BS (2010). Quantitative differences in intestinal Faecalibacterium prausnitzii in obese Indian children. *British Journal of Nutrition* **103:** 335-338.

Bellia A, Giardina E, Lauro D, Tesauro M, Di Fede G, Cusumano G *et al.* (2009). "The Linosa Study": Epidemiological and heritability data of the metabolic syndrome in a Caucasian genetic isolate. *Nutrition, Metabolism and Cardiovascular Diseases* **19:** 455-461.

Benson AK, Kelly SA, Legge R, Ma F, Low SJ, Kim J *et al.* (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proceedings of the National Academy of Sciences* **107:** 18933-18938.

Blainey PC, Milla CE, Cornfield DN, Quake SR (2012). Quantitative analysis of the human airway microbial ecology reveals a pervasive signature for cystic fibrosis. *Science translational medicine* **4:** 153ra130.

Booth ML, Ainsworth BE, Pratt M, Ekelund U, Yngve A, Sallis JF *et al.* (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine and Science in Sports and Exercise* **195:** 3508-1381.

Candela M, Perna F, Carnevali P, Vitali B, Ciati R, Gionchetti P *et al.* (2008). Interaction of probiotic *Lactobacillus* and *Bifidobacterium* strains with human intestinal epithelial cells: adhesion properties, competition against enteropathogens and modulation of IL-8 production. *International Journal of Food Microbiology* **125:** 286-292.

Cani PD, Bibiloni R, Knauf C, Waget A, Neyrinck AM, Delzenne NM *et al.* (2008). Changes in gut microbiota control metabolic endotoxemia-induced inflammation in high-fat diet–induced obesity and diabetes in mice. *Diabetes* **57:** 1470-1481.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7:** 335-336.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ *et al.* (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences* **108:** 4516-4522.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N *et al.* (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME journal* **6:** 1621-1624.

Cebra JJ (1999). Influences of microbiota on intestinal immune system development. *The American Journal of Clinical Nutrition* **69:** 1046s-1051s.

Cha H-R, Chang S-Y, Chang J-H, Kim J-O, Yang J-Y, Kim C-H *et al.* (2010). Downregulation of Th17 cells in the small intestine by disruption of gut flora in the absence of retinoic acid. *Journal of Immunology* **184:** 6799-6806.

Charlson ES, Chen J, Custers-Allen R, Bittinger K, Li H, Sinha R *et al.* (2010). Disordered microbial communities in the upper respiratory tract of cigarette smokers. *PloS one* **5:** e15216.

Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A *et al.* (2011). Topographical continuity of bacterial populations in the healthy human respiratory tract. *American journal of respiratory and critical care medicine* **184:** 957-963.

Chen W, Brehm JM, Manichaikul A, Cho MH, Boutaoui N, Yan Q *et al.* (2015). A Genome-Wide Association Study of Chronic Obstructive Pulmonary Disease in Hispanics. *Annals of the American Thoracic Society* **12:** 340-348.

Cho I and Blaser MJ (2012). The human microbiome: at the interface of health and disease. *Nature Reviews Genetics* **13:** 260-270.

Cho MH, McDonald M-LN, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP *et al.* (2014). Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med* **2:** 214-225.

Claesson M, Jeffery I, Conde S, Power S, O'Connor E, Cusack S *et al.* (2012). Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488:** 178-184.

Clarke KR (1993). Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology* **18:** 117-143.

Cui Z, Zhou Y, Li H, Zhang Y, Zhang S, Tang S *et al.* (2012). Complex sputum microbial composition in patients with pulmonary tuberculosis. *BMC Microbiol* **12:** 276.

Daniels T, Rogers G, Stressmann F, van der Gast C, Bruce K, Jones G *et al.* (2013). Impact of

antibiotic treatment for pulmonary exacerbations on bacterial diversity in cystic fibrosis. *Journal of Cystic Fibrosis* **12:** 22-28.

De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet J, Massart S *et al.* (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences* **107:** 14691-14696.

Delwiche E, Pestka J, Tortorello M (1985). The veillonellae: gram-negative cocci with a unique physiology. *Annual Reviews in Microbiology* **39:** 175-193.

Delzenne NM, Neyrinck AM, Bäckhed F, Cani PD (2011). Targeting gut microbiota in obesity: effects of prebiotics and probiotics. *Nature Reviews Endocrinology* **7:** 639-646.

Demling RH (2008). Smoke inhalation lung injury: an update. *Eplasty* **8:** e27.

Dethlefsen L and Relman DA (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proceedings of the National Academy of Sciences* **108:** 4554-4561.

Ege MJ, Mayer M, Normand A-C, Genuneit J, Cookson WO, Braun-Fahrländer C *et al.* (2011). Exposure to environmental microorganisms and childhood asthma. *New England Journal of Medicine* **364:** 701-709.

Erb-Downward JR, Thompson DL, Han MK, Freeman CM, McCloskey L, Schmidt LA *et al.* (2011). Analysis of the lung microbiome in the "healthy" smoker and in COPD. *PloS one* **6:** e16384.

Esposito K, Marfella R, Ciotola M, Di Palo C, Giugliano F, Giugliano G *et al.* (2004). Effect of a Mediterranean-style diet on endothelial dysfunction and markers of vascular inflammation in the metabolic syndrome: a randomized trial. *The Journal of the American Medical Association* **292:** 1440-1446.

Everard A, Belzer C, Geurts L, Ouwerkerk JP, Druart C, Bindels LB *et al.* (2013). Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. *Proceedings of the National Academy of Sciences* **110:** 9066-9071.

Fink DL and Geme JW (2006). The Genus Haemophilus. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (eds). *The Prokaryotes***:** 1034-1061.

Ford ES (2005). Prevalence of the metabolic syndrome defined by the International Diabetes

Federation among adults in the US. *Diabetes Care* **28:** 2745-2749.

Foster G, Higgins R, Leclair D, Korczak BM, Mikaelian I, Patterson IAP *et al.* (2011). Proposal of Bisgaardia hudsonensis gen. nov., sp. nov. and an additional genomospecies, isolated from seals, as new members of the family Pasteurellaceae. *International journal of systematic and evolutionary microbiology* **61:** 3016-3022.

Fujimura KE, Johnson CC, Ownby DR, Cox MJ, Brodie EL, Havstad SL *et al.* (2010). Man's best friend? The effect of pet ownership on house dust microbial communities. *The Journal of Allergy and Clinical Immunology* **126:** 410-412.

Gill SR, Pop M, DeBoy RT, Eckburg PB, Turnbaugh PJ, Samuel BS *et al.* (2006). Metagenomic analysis of the human distal gut microbiome. *Science* **312:** 1355-1359.

Gombojav B, Song Y-M, Lee K, Yang S, Kho M, Hwang Y-C *et al.* (2013). The Healthy Twin Study, Korea updates: resources for omics and genome epidemiology studies. *Twin Research and Human Genetics* **16:** 241-245.

Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R *et al.* (2014). Human genetics shape the gut microbiome. *Cell* **159:** 789-799.

Grundy SM, Cleeman JI, Daniels SR, Donato KA, Eckel RH, Franklin BA *et al.* (2005). Diagnosis and management of the metabolic syndrome an American Heart Association/National Heart, Lung, and Blood Institute scientific statement. *Circulation* **112:** 2735-2752.

Grundy SM (2008). Metabolic syndrome pandemic. *Arteriosclerosis, Thrombosis, and Vascular Biology* **28:** 629-636.

Guardiola M, Alvaro A, Vallvé J, Rosales R, Solà R, Girona J *et al.* (2012). APOA5 gene expression in the human intestinal tissue and its response to in vitro exposure to fatty acid and fibrate. *Nutrition, Metabolism and Cardiovascular Diseases* **22:** 756-762.

Haberman Y, Tickle TL, Dexheimer PJ, Kim M-O, Tang D, Karns R *et al.* (2014). Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *The Journal of Clinical Investigation* **124:** 3617.

Hampton TH, Green DM, Cutting GR, Morrison HG, Sogin ML, Gifford AH *et al.* (2014). The microbiome in pediatric cystic fibrosis patients: the role of shared environment suggests a window of intervention. *Microbiome* **2:** 14.

Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO (1991). The Fagerström test for nicotine dependence: a revision of the Fagerstrom Tolerance Questionnaire. *British journal of addiction* **86:** 1119-1127.

Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C *et al.* (2010). Disordered microbial communities in asthmatic airways. *PloS one* **5:** e8578.

Hong K-W, Jeong SW, Chung M, Cho SB (2014). Association between expression quantitative trait loci and metabolic traits in two Korean populations. *PLoS ONE* **9:** e114128.

Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. *Nature* **486:** 207-214.

Huse S, Ye Y, Zhou Y, Fodor A (2012). A core human microbiome as viewed through 16S rRNA sequence clusters. *PLoS ONE* **7:** e34242.

Kadooka Y, Sato M, Imaizumi K, Ogawa A, Ikuyama K, Akai Y *et al.* (2010). Regulation of abdominal adiposity by probiotics (Lactobacillus gasseri SBT2055) in adults with obese tendencies in a randomized controlled trial. *European Journal of Clinical Nutrition* **64:** 636-643.

Karlsson F, Fåk F, Nookaew I, Tremaroli V, Fagerberg B, Petranovic D *et al.* (2012). Symptomatic atherosclerosis is associated with an altered gut metagenome. *Nature Communications* **3:** 1245.

Karlsson FH, Ussery DW, Nielsen J, Nookaew I (2011). A closer look at *bacteroides*: phylogenetic relationship and genomic implications of a life in the human gut. *Microbial Ecology* **61:** 473-485.

Kaufman L and Rousseeuw PJ (2009). *Finding groups in data: an introduction to cluster analysis*, vol. 344. John Wiley & Sons: New York.

Kaur J (2014). A comprehensive review on metabolic syndrome. *Cardiology Research and Practice* **2014:** 943162.

Koren O, Knights D, Gonzalez A, Waldron L, Segata N, Knight R *et al.* (2013). A guide to enterotypes across the human body: meta-analysis of microbial community structures in human microbiome datasets. *PLoS computational biology* **9:** e1002863.

Kovacs A, Ben-Jacob N, Tayem H, Halperin E, Iraqi FA, Gophna U (2011). Genotype is a stronger determinant than sex of the mouse gut microbiota. *Microbial Ecology* **61:** 423-428.

Kraja AT, Vaidya D, Pankow JS, Goodarzi MO, Assimes TL, Kullo IJ *et al.* (2011). A bivariate genome-wide approach to metabolic syndrome STAMPEED consortium. *Diabetes* **60:** 1329-1339.

Løkke A, Lange P, Scharling H, Fabricius P, Vestbo J (2006). Developing COPD: a 25 year follow up study of the general population. *Thorax* **61:** 935-939.

Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA *et al.* (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology* **31:** 814-821.

Larsen N, Vogensen FK, van den Berg FW, Nielsen DS, Andreasen AS, Pedersen BK *et al.* (2010). Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS ONE* **5:** e9085.

Lee SY, Park HS, Kim DJ, Han JH, Kim SM, Cho GJ *et al.* (2007). Appropriate waist circumference cutoff points for central obesity in Korean adults. *Diabetes Research and Clinical Practice* **75:** 72-80.

Ley RE, Turnbaugh PJ, Klein S, Gordon JI (2006). Microbial ecology: human gut microbes associated with obesity. *Nature* **444:** 1022-1023.

Lieberoth S, Backer V, Kyvik KO, Skadhauge LR, Tolstrup JS, Grønbæk M *et al.* (2012). Intake of alcohol and risk of adult-onset asthma. *Respir Med* **106:** 184-188.

Lim S, Shin H, Song JH, Kwak SH, Kang SM, Yoon JW *et al.* (2011). Increasing prevalence of metabolic syndrome in Korea The Korean National Health and Nutrition Examination Survey for 1998–2007. *Diabetes Care* **34:** 1323-1328.

Lin H-F, Boden-Albala B, Juo S, Park N, Rundek T, Sacco R (2005). Heritabilities of the metabolic syndrome and its components in the Northern Manhattan Family Study. *Diabetologia* **48:** 2006-2012.

Lindberg A, Jonsson A-C, Ronmark E, Lundgren R, Larsson L-G, Lundback B (2005). Ten-year cumulative incidence of COPD and risk factors for incident disease in a symptomatic cohort. *CHEST Journal* **127:** 1544-1552.

Maccarrone M, Gasperi V, Catani MV, Diep TA, Dainese E, Hansen HS *et al.* (2010). The endocannabinoid system and its relevance for nutrition. *Annual Review of Nutrition* **30:** 423-440.

Maggini S, Wintergerst E, Beveridge S, Hornig D (2007). Selected vitamins and trace elements support immune function by strengthening epithelial barriers and cellular and humoral immune responses. *The British Journal of Nutrition* **98:** S29-S35.

Marsland BJ and Gollwitzer ES (2014). Host-microorganism interactions in lung diseases. *Nature Reviews Immunology* **14:** 827-835.

McKnite AM, Perez-Munoz ME, Lu L, Williams EG, Brewer S, Andreux PA *et al.* (2012). Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. *PLoS ONE* **7:** e39191.

Million M, Maraninchi M, Henry M, Armougom F, Richet H, Carrieri P *et al.* (2011). Obesity-associated gut microbiota is enriched in Lactobacillus reuteri and depleted in Bifidobacterium animalis and Methanobrevibacter smithii. *International Journal of Obesity* **36:** 817-825.

Million M, Angelakis E, Maraninchi M, Henry M, Giorgi R, Valero R *et al.* (2013). Correlation between body mass index and gut concentrations of Lactobacillus reuteri, Bifidobacterium animalis, Methanobrevibacter smithii and Escherichia coli. *International Journal of Obesity*.

Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV *et al.* (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* **13:** R79.

Morgan XC and Huttenhower C (2014). Meta'omic analytic techniques for studying the intestinal microbiome. *Gastroenterology* **146:** 1437-1448. e1431.

Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL *et al.* (2013). Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *American journal of respiratory and critical care medicine* **187:** 1067-1075.

Nasidze I, Li J, Quinque D, Tang K, Stoneking M (2009). Global diversity in the human salivary microbiome. *Genome Res* **19:** 636-643.

Nicolle LE (2002). Resistant pathogens in urinary tract infections. *J Am Geriatr Soc* **50:** 230-235.

O'Hara CM, Brenner FW, Miller JM (2000). Classification, identification, and clinical significance of Proteus, Providencia, and Morganella. *Clin Microbiol Rev* **13:** 534-546.

O'Brien PJ, Alborn WE, Sloan JH, Ulmer M, Boodhoo A, Knierman MD *et al.* (2005). The novel apolipoprotein A5 is present in human serum, is associated with VLDL, HDL, and chylomicrons, and circulates at very low concentrations compared with other apolipoproteins. *Clinical Chemistry* **51:** 351-359.

Ou J, Carbonero F, Zoetendal E, DeLany J, Wang M, Newton K *et al.* (2013). Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans. *The American Journal of Clinical Nutrition* **98:** 111-120.

Pennacchio LA, Olivier M, Hubacek JA, Cohen JC, Cox DR, Fruchart J-C *et al.* (2001). An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science* **294:** 169-173.

Pflughoeft KJ and Versalovic J (2012). Human microbiome in health and disease. *Annual Review of Pathology: Mechanisms of Disease* **7:** 99-122.

Philipp R, Ateequr R, Sven K, Robert H, Stephan JO, Stefan S *et al.* (2011). Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. *Proceedings of the National Academy of Sciences* **108:** 19030-19035.

Piipari R, Jaakkola J, Jaakkola N, Jaakkola M (2004). Smoking and asthma in adults. *Eur Respir J* **24:** 734-739.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81:** 559-575.

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C *et al.* (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464:** 59-65.

Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F *et al.* (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490:** 55-60.

Rajilić-Stojanović M, Heilig H, Tims S, Zoetendal E, de Vos W (2012). Long-term monitoring of the human intestinal microbiota composition. *Environmental Microbiology* **15:** 1146–1159.

Reilly KH, Gu D, Duan X, Wu X, Chen C-S, Huang J *et al.* (2008). Risk factors for chronic

obstructive pulmonary disease mortality in Chinese adults. *Am J Epidemiol* **167:** 998-1004.

Ribalta J, Figuera L, Fernández-Ballart J, Vilella E, Cabezas MC, Masana L *et al.* (2002). Newly identified apolipoprotein AV gene predisposes to high plasma triglycerides in familial combined hyperlipidemia. *Clinical Chemistry* **48:** 1597-1600.

Roager H, Licht T, Poulsen S, Larsen T, Bahl M (2014). Microbial enterotypes, inferred by the *Prevotella*-to-*Bacteroides* ratio, remained stable during a 6-month randomized controlled diet intervention with the new nordic diet. *Applied and Environmental Microbiology* **80:** 1142-1149.

Rogers G, Carroll M, Serisier D, Hockey P, Jones G, Kehagia V *et al.* (2006). Use of 16S rRNA gene profiling by terminal restriction fragment length polymorphism analysis to compare bacterial communities in sputum and mouthwash samples from patients with cystic fibrosis. *Journal of clinical microbiology* **44:** 2601-2604.

Rogosa M (1964). The genus Veillonella I. General cultural, ecological, and biochemical considerations. *Journal of bacteriology* **87:** 162-170.

Samuel BS, Shaito A, Motoike T (2008). 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* **105:** 16767-16772.

Sanz Y, Olivares M, Moya-Pérez Á, Agostoni C (2014). Understanding the role of gut microbiome in metabolic disease risk. *Pediatric Research* **77:** 236–244.

Schaap FG, Rensen PC, Voshol PJ, Vrins C, van der Vliet HN, Chamuleau RA *et al.* (2004). ApoAV reduces plasma triglycerides by inhibiting very low density lipoprotein-triglyceride (VLDL-TG) production and stimulating lipoprotein lipase-mediated VLDL-TG hydrolysis. *Journal of Biological Chemistry* **279:** 27941-27947.

Scher J, Sczesnak A, Longman R, Segata N, Ubeda C, Bielski C *et al.* (2013). Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. *eLife* **2:** e01202.

Segal LN, Alekseyenko AV, Clemente JC, Kulkarni R, Wu B, Chen H *et al.* (2013). Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. *Microbiome* **1:** 19.

Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett W *et al.* (2011). Metagenomic biomarker discovery and explanation. *Genome Biology* **12:** R60.

Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods* **9:** 811-814.

Sethi S (2010). Infection as a comorbidity of COPD. *Eur Respir J* **35:** 1209-1215.

Shah HN and Collins DM (1990). NOTES: *Prevotella*, a new genus to include *Bacteroides melaninogenicus* and related species formerly classified in the genus *Bacteroides*. *International Journal of Systematic Bacteriology* **40:** 205-208.

Smith C and Hansch C (2000). The relative toxicity of compounds in mainstream cigarette smoke condensate. *Food and Chemical Toxicology* **38:** 637-646.

Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T (2011). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27:** 431-432.

Sommer F and Bäckhed F (2013). The gut microbiota—masters of host development and physiology. *Nature Reviews Microbiology* **11:** 227-238.

Sorensen L (1959). Degradation of uric acid in man. *Metabolism: Clinical and Experimental* **8:** 687-703.

Stämpfli MR and Anderson GP (2009). How cigarette smoke skews immune responses to promote infection, lung disease and cancer. *Nature Reviews Immunology* **9:** 377-384.

Stewart M, Savarino V, Slavin J (2009). Assessment of dietary fiber fermentation: effect of *Lactobacillus reuteri* and reproducibility of short-chain fatty acid concentrations. *Molecular Nutrition & Food Research* **53:** S114-S120.

Stokell JR, Gharaibeh RZ, Hamp TJ, Zapata MJ, Fodor AA, Steck TR (2015). Analysis of changes in diversity and abundance of the microbial community in a cystic fibrosis patient over a multiyear period. *Journal of Clinical Microbiology* **53:** 237-247.

Sung J, Cho S-I, Lee K, Ha M, Choi E-Y, Choi J-S *et al.* (2006). Healthy Twin: a twin-family study of Korea—protocols and current status. *Twin Research and Human Genetics* **9:** 844-848.

Sung J, Lee K, Song Y-M (2009). Heritabilities of the metabolic syndrome phenotypes and related factors in Korean twins. *The Journal of Clinical Endocrinology & Metabolism* **94:** 4946-4952.

Tagliabue A and Elli M (2013). The role of gut microbiota in human obesity: Recent findings and future perspectives. *Nutrition, Metabolism and Cardiovascular Diseases* **23:** 160-168.

Tan A, Sun J, Xia N, Qin X, Hu Y, Zhang S *et al.* (2012). A genome-wide association and gene–environment interaction study for serum triglycerides levels in a healthy Chinese male population. *Human Molecular Genetics* **21:** 1658-1664.

Tong M, McHardy I, Ruegger P, Goudarzi M, Kashyap PC, Haritunians T *et al.* (2014). Reprograming of gut microbiome energy metabolism by the FUT2 Crohn's disease risk polymorphism. *The ISME journal* **8:** 2193-2206.

Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE *et al.* (2011). Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet* **43:** 887-892.

Turnbaugh P, Hamady M, Yatsunenko T, Cantarel B, Duncan A, Ley R *et al.* (2009). A core gut microbiome in obese and lean twins. *Nature* **457:** 480-484.

Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444:** 1027-1131.

Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI (2007). The human microbiome project. *Nature* **449:** 804-810.

van der Gast CJ, Walker AW, Stressmann FA, Rogers GB, Scott P, Daniels TW *et al.* (2011). Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *The ISME journal* **5:** 780-791.

van der Vliet HN, Schaap FG, Levels JH, Ottenhoff R, Looije N, Wesseling JG *et al.* (2002). Adenoviral overexpression of apolipoprotein AV reduces serum levels of triglycerides and cholesterol in mice. *Biochemical and Biophysical Research Communications* **295:** 1156-1159.

Vijay-Kumar M, Aitken JD, Carvalho FA, Cullender TC, Mwangi S, Srinivasan S *et al.* (2010). Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328:** 228-231.

Walter J (2008). Ecological role of lactobacilli in the gastrointestinal tract: implications for fundamental and biomedical research. *Applied and Environmental Microbiology* **74:** 4985-4996.

Wang J, Tang H, Zhang C, Zhao Y, Derrien M, Rocher E *et al.* (2015). Modulation of gut microbiota during probiotic-mediated attenuation of metabolic syndrome in high fat diet-fed mice. *The ISME journal* **9:** 1-15.

Werner T, Wagner S, Martínez I, Walter J, Chang J-S, Clavel T *et al.* (2011). Depletion of luminal iron alters the gut microbiota and prevents Crohn's disease-like ileitis. *Gut* **60:** 325-333.

Wilson PW, D'Agostino RB, Parise H, Sullivan L, Meigs JB (2005). Metabolic syndrome as a precursor of cardiovascular disease and type 2 diabetes mellitus. *Circulation* **112:** 3066-3072.

Wu G, Chen J, Hoffmann C, Bittinger K, Chen Y-Y, Keilbaugh S *et al.* (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334:** 105-108.

Wu J, Liu W, He L, Huang F, Chen J, Cui P *et al.* (2013). Sputum microbiota associated with new, recurrent and treatment failure tuberculosis. *PloS one* **8:** e83445.

Wu X, Ma C, Han L, Nawaz M, Gao F, Zhang X *et al.* (2010). Molecular characterisation of the faecal microbiota in patients with type II diabetes. *Current Microbiology* **61:** 69-78.

Yamada Y, Ichihara S, Kato K, Yoshida T, Yokoi K, Matsuo H *et al.* (2008). Genetic risk for metabolic syndrome: examination of candidate gene polymorphisms related to lipid metabolism in Japanese people. *Journal of Medical Genetics* **45:** 22-28.

Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M *et al.* (2012). Human gut microbiome viewed across age and geography. *Nature* **486:** 222-227.

Zaura E, Keijser BJ, Huse SM, Crielaard W (2009). Defining the healthy "core microbiome" of oral microbial communities. *BMC Microbiol* **9:** 259.

Zhang C, Zhang M, Wang S, Han R, Cao Y, Hua W *et al.* (2009a). Interactions between gut microbiota, host genetics and diet relevant to development of metabolic syndromes in mice. *The ISME Journal* **4:** 232-241.

Zhang H, DiBaise JK, Zuccolo A, Kudrna D, Braidotti M, Yu Y *et al.* (2009b). Human gut microbiota in obesity and after gastric bypass. *Proceedings of the National Academy of Sciences* **106:** 2365-2370.

Zhang LS, Xu M, Yang Q, Ryan RO, Howles PN, Tso P (2015). Apolipoprotein AV deficiency

enhances chylomicron production in lymph fistula mice. *American Journal of Physiology-Gastrointestinal and Liver Physiology***:** ajpgi. 00339.02014.

Zhang X, Shen D, Fang Z, Jie Z, Qiu X, Zhang C *et al.* (2013). Human gut microbiota changes reveal the progression of glucose intolerance. *PLoS ONE* **8:** e71108.

Zhou Y, Mihindukulasuriya KA, Gao H, La Rosa PS, Wylie KM, Martin JC *et al.* (2014). Exploration of bacterial community classes in major human habitats. *Genome Biology* **15:** R66.

Zhu L, Baker S, Gill C, Liu W, Alkhouri R, Baker R *et al.* (2013). Characterization of gut microbiomes in nonalcoholic steatohepatitis (NASH) patients: a connection between endogenous alcohol and NASH. *Hepatology (Baltimore, Md)* **57:** 601-609.

Zimmermann M, Chassard C, Rohner F, N'Goran E, Nindjin C, Dostal A *et al.* (2010). The effects of iron fortification on the gut microbiota in African children: a randomized controlled trial in Cote d'Ivoire. *The American Journal of Clinical Nutrition* **92:** 1406-1415.

Zupancic ML, Cantarel BL, Liu Z, Drabek EF, Ryan KA, Cirimotich S *et al.* (2012). Analysis of the gut microbiota in the old order Amish and its relation to the metabolic syndrome. *PLoS ONE* **7:** e43052.

Appendices

Appendix A. Heritability of the gut microbiota.

Taxon	H2r	p.H2r	se.H2r	FDR.H2r	C2	p.C2	se.C2	FDR.C2	p.sex	p.age	p.MetS	Group
k__Archaea	0.2222	0.0000	0.0576	0.0001	0.0000	NA	NA	NA	0.7592	0.0332	0.1594	
p__Euryarchaeota	0.2222	0.0000	0.0576	0.0001	0.0000	NA	NA	NA	0.7592	0.0332	0.1594	
p__Euryarchaeota|o__Methanobacteriales	0.2254	0.0000	0.0577	0.0001	0.0000	NA	NA	NA	0.7327	0.0299	0.1590	
p__Euryarchaeota|f__Methanobacteriaceae	0.2254	0.0000	0.0577	0.0001	0.0000	NA	NA	NA	0.7327	0.0299	0.1590	
p__Euryarchaeota|g__Methanobrevibacter	0.2085	0.0001	0.0595	0.0004	0.0000	NA	NA	NA	0.5688	0.1840	0.3873	MetS
k__Bacteria	0.2451	0.0000	0.0563	0.0000	0.0000	NA	NA	NA	0.4468	0.1264	0.3435	
p__Actinobacteria	0.4569	0.0000	0.0548	0.0000	0.0000	NA	NA	NA	0.2051	0.0000	0.1055	Healthy
p__Actinobacteria|c__Actinobacteria	0.4112	0.0000	0.0561	0.0000	0.0000	NA	NA	NA	0.2319	0.0000	0.1294	Healthy
p__Actinobacteria|o__Bifidobacteriales	0.3724	0.0002	0.0953	0.0005	0.0286	0.3464	0.0737	0.4606	0.1456	0.0000	0.3518	Healthy
p__Actinobacteria|f__Bifidobacteriaceae	0.3724	0.0002	0.0953	0.0005	0.0286	0.3464	0.0737	0.4606	0.1456	0.0000	0.3518	Healthy
p__Actinobacteria|g__Bifidobacterium	0.3740	0.0001	0.0941	0.0005	0.0198	0.3911	0.0727	0.4620	0.2077	0.0000	0.3585	Healthy
p__Bacteroidetes	0.2520	0.0084	0.1005	0.0164	0.0455	0.2962	0.0874	0.4606	0.3769	0.8094	0.5422	
p__Bacteroidetes|o__Bacteroidales	0.2527	0.0083	0.1005	0.0164	0.0451	0.2978	0.0874	0.4606	0.3785	0.8079	0.5363	
p__Bacteroidetes|f__Bacteroidaceae	0.2383	0.0000	0.0595	0.0001	0.0000	NA	NA	NA	0.0000	0.0053	0.9052	
p__Bacteroidetes|g__Bacteroides	0.2381	0.0000	0.0595	0.0001	0.0000	NA	NA	NA	0.0000	0.0048	0.9243	
p__Bacteroidetes|f__Porphyromonadaceae	0.1527	0.0027	0.0583	0.0061	0.0000	NA	NA	NA	0.0172	0.1113	0.0596	Healthy
p__Bacteroidetes|g__Parabacteroides	0.1567	0.0024	0.0591	0.0059	0.0000	NA	NA	NA	0.0165	0.1186	0.0772	Healthy
p__Bacteroidetes|f__Prevotellaceae	0.1487	0.0628	0.0959	0.1007	0.0737	0.1739	0.0812	0.4606	0.0028	0.0123	0.9730	

Taxon	H2r	p.H2r	se.H2r	FDR.H2r	C2	p.C2	se.C2	FDR.C2	p.sex	p.age	p.MetS	Group
p__Bacteroidetes|g__Prevotella	0.1545	0.0567	0.0962	0.0927	0.0714	0.1826	0.0814	0.4606	0.0034	0.0145	0.9815	
p__Bacteroidetes|f__Rikenellaceae	0.1299	0.0804	0.0911	0.1139	0.0704	0.1678	0.0759	0.4606	0.0000	0.1564	0.0051	Healthy
p__Bacteroidetes|f__[Odoribacteraceae]	0.1918	0.0276	0.0975	0.0469	0.0580	0.2241	0.0787	0.4606	0.0498	0.1234	0.0029	Healthy
p__Bacteroidetes|g__Butyricimonas	0.1438	0.0906	0.1058	0.1262	0.0640	0.2124	0.0828	0.4606	0.3988	0.2535	0.0190	
p__Bacteroidetes|g__Odoribacter	0.1916	0.0002	0.0575	0.0006	0.0000	NA	NA	NA	0.0020	0.2504	0.0350	Healthy
p__Bacteroidetes|f__[Paraprevotellaceae]	0.0833	0.2201	0.1067	0.2563	0.0964	0.1196	0.0863	0.4606	0.3472	0.0009	0.0667	Healthy
p__Bacteroidetes|g__Paraprevotella	0.1412	0.0043	0.0561	0.0091	0.0000	NA	NA	NA	0.6257	0.0565	0.0589	Healthy
p__Bacteroidetes|g__[Prevotella]	0.0439	0.3435	0.1086	0.3605	0.0723	0.1949	0.0880	0.4606	0.1594	0.0008	0.9142	
p__Cyanobacteria	0.2748	0.0000	0.0654	0.0000	0.0000	NA	NA	NA	0.0792	0.0098	0.2204	
p__Firmicutes	0.1019	0.1678	0.1052	0.2067	0.0718	0.1937	0.0859	0.4606	0.1613	0.1159	0.5852	
p__Firmicutes|c__Bacilli	0.3528	0.0000	0.0569	0.0000	0.0000	NA	NA	NA	0.9764	0.0019	0.1743	
p__Firmicutes|o__Lactobacillales	0.3282	0.0000	0.0570	0.0000	0.0000	NA	NA	NA	0.8384	0.0015	0.1547	
p__Firmicutes|f__Lactobacillaceae	0.1708	0.0010	0.0591	0.0029	0.0000	NA	NA	NA	0.8536	0.0023	0.0835	MetS
p__Firmicutes|g__Lactobacillus	0.1481	0.0040	0.0594	0.0086	0.0000	NA	NA	NA	0.8830	0.0037	0.0882	MetS
p__Firmicutes|f__Leuconostocaceae	0.1365	0.0780	0.0951	0.1123	0.1437	0.0293	0.0812	0.4493	0.3515	0.0004	0.0993	
p__Firmicutes|f__Streptococcaceae	0.2909	0.0017	0.0937	0.0043	0.0412	0.2829	0.0737	0.4606	0.8866	0.0794	0.9084	
p__Firmicutes|g__Streptococcus	0.3186	0.0008	0.0949	0.0025	0.0242	0.3697	0.0739	0.4620	0.8755	0.2575	0.9764	
p__Firmicutes|c__Clostridia	0.0719	0.2454	0.1040	0.2818	0.0923	0.1287	0.0852	0.4606	0.3155	0.0189	0.4784	
p__Firmicutes|o__Clostridiales	0.0693	0.2530	0.1040	0.2858	0.0945	0.1230	0.0853	0.4606	0.3129	0.0185	0.4730	
p__Firmicutes|f__Christensenellaceae	0.3064	0.0000	0.0593	0.0000	0.0000	NA	NA	NA	0.0018	0.0195	0.0310	Healthy
p__Firmicutes|f__Clostridiaceae	0.3365	0.0000	0.0572	0.0000	0.0000	NA	NA	NA	0.0679	0.0079	0.7328	
p__Firmicutes|g__Clostridium	0.1427	0.0764	0.0973	0.1120	0.0470	0.2629	0.0765	0.4606	0.0088	0.6926	0.6667	

Taxon	H2r	p.H2r	se.H2r	FDR.H2r	C2	p.C2	se.C2	FDR.C2	p.sex	p.age	p.MetS	Group
p__Firmicutes|f__Lachnospiraceae	0.1542	0.0014	0.0564	0.0037	0.0000	NA	NA	NA	0.2963	0.3847	0.9963	
p__Firmicutes|g__Anaerostipes	0.0120	0.4482	0.0919	0.4535	0.0543	0.2288	0.0751	0.4606	0.6073	0.8601	0.0984	
p__Firmicutes|g__Blautia	0.1946	0.0235	0.0969	0.0408	0.0365	0.3030	0.0719	0.4606	0.7673	0.0000	0.5957	
p__Firmicutes|g__Coprococcus	0.1628	0.0011	0.0571	0.0029	0.0000	NA	NA	NA	0.1723	0.1506	0.7319	
p__Firmicutes|g__Dorea	0.0278	0.3933	0.1025	0.4077	0.1635	0.0200	0.0848	0.4493	0.0299	0.0018	0.3575	MetS
p__Firmicutes|g__Lachnospira	0.1533	0.0025	0.0584	0.0059	0.0000	NA	NA	NA	0.4834	0.1164	0.2724	
p__Firmicutes|g__Roseburia	0.0000	0.5000	NA	0.5000	0.0679	0.1574	0.0418	0.4606	0.0780	0.4168	0.4604	
p__Firmicutes|g__[Ruminococcus]	0.2084	0.0193	0.0984	0.0342	0.1036	0.0884	0.0809	0.4606	0.0002	0.0000	0.8004	
p__Firmicutes|f__Ruminococcaceae	0.0546	0.2915	0.0992	0.3137	0.1731	0.0149	0.0851	0.4493	0.0034	0.1575	0.3533	
p__Firmicutes|g__Faecalibacterium	0.2631	0.0000	0.0572	0.0000	0.0000	NA	NA	NA	0.1092	0.1204	0.6208	
p__Firmicutes|g__Oscillospira	0.1434	0.0750	0.0969	0.1120	0.0088	0.4534	0.0751	0.4733	0.0000	0.1641	0.0566	
p__Firmicutes|g__Ruminococcus	0.0562	0.1575	0.0581	0.1984	0.0000	NA	NA	NA	0.0010	0.5001	0.6916	
p__Firmicutes|f__Veillonellaceae	0.1343	0.0085	0.0589	0.0164	0.0000	NA	NA	NA	0.0000	0.0172	0.0773	
p__Firmicutes|g__Acidaminococcus	0.1312	0.0067	0.0562	0.0140	0.0000	NA	NA	NA	0.0156	0.2438	0.9763	
p__Firmicutes|g__Dialister	0.1636	0.0028	0.0628	0.0064	0.0000	NA	NA	NA	0.3238	0.0089	0.2050	
p__Firmicutes|g__Megamonas	0.2177	0.0001	0.0633	0.0005	0.0000	NA	NA	NA	0.0007	0.0023	0.5567	
p__Firmicutes|g__Megasphaera	0.1375	0.0089	0.0619	0.0169	0.0000	NA	NA	NA	0.0001	0.4755	0.0332	
p__Firmicutes|g__Phascolarctobacterium	0.3549	0.0002	0.0929	0.0006	0.0045	0.4733	0.0669	0.4733	0.0270	0.4549	0.3422	
p__Firmicutes|g__Succiniclasticum	0.0139	0.3983	0.0543	0.4079	0.0000	NA	NA	NA	0.5272	0.5507	0.8409	
p__Firmicutes|g__Veillonella	0.2168	0.0171	0.0998	0.0310	0.0531	0.2745	0.0911	0.4606	0.3765	0.0001	0.6574	
p__Firmicutes|f__Erysipelotrichaceae	0.1402	0.0744	0.0955	0.1120	0.0843	0.1287	0.0781	0.4606	0.0127	0.2363	0.9147	

Taxon	H2r	p.H2r	se.H2r	FDR.H2r	C2	p.C2	se.C2	FDR.C2	p.sex	p.age	p.MetS	Group
p__Firmicutes|g__Catenibacterium	0.1208	0.1216	0.1025	0.1656	0.0212	0.3917	0.0781	0.4620	0.0793	0.5195	0.9312	
p__Firmicutes|g__[Eubacterium]	0.2831	0.0000	0.0592	0.0000	0.0000	NA	NA	NA	0.0053	0.5167	0.6889	
p__Fusobacteria|o__Fusobacteriales	0.1037	0.1587	0.1024	0.1984	0.0581	0.2403	0.0847	0.4606	0.6476	0.6526	0.3199	
p__Fusobacteria|f__Fusobacteriaceae	0.0946	0.1734	0.0997	0.2105	0.0908	0.1265	0.0827	0.4606	0.4864	0.5263	0.6091	
p__Proteobacteria	0.1385	0.0025	0.0527	0.0059	0.0000	NA	NA	NA	0.6985	0.0000	0.2652	
p__Proteobacteria|c__Betaproteobacteria	0.0731	0.2110	0.0899	0.2491	0.0290	0.3364	0.0702	0.4606	0.8888	0.8048	0.0634	MetS
p__Proteobacteria|o__Burkholderiales	0.0759	0.2098	0.0929	0.2491	0.0272	0.3505	0.0721	0.4606	0.9092	0.7426	0.0605	MetS
p__Proteobacteria|f__Alcaligenaceae	0.1344	0.0761	0.0912	0.1120	0.0064	0.4635	0.0706	0.4733	0.6601	0.7359	0.0387	MetS
p__Proteobacteria|g__Sutterella	0.1361	0.0735	0.0912	0.1120	0.0051	0.4710	0.0705	0.4733	0.6684	0.7389	0.0383	MetS
p__Proteobacteria|c__Deltaproteobacteria	0.0987	0.1569	0.0971	0.1984	0.0088	0.4547	0.0777	0.4733	0.5911	0.0234	0.0664	
p__Proteobacteria|o__Desulfovibrionales	0.1010	0.1506	0.0967	0.1969	0.0063	0.4676	0.0776	0.4733	0.5962	0.0211	0.0696	
p__Proteobacteria|f__Desulfovibrionaceae	0.1012	0.1500	0.0967	0.1969	0.0061	0.4685	0.0776	0.4733	0.5958	0.0211	0.0697	
p__Proteobacteria|g__Bilophila	0.0501	0.2982	0.0944	0.3168	0.0360	0.3222	0.0792	0.4606	0.1134	0.0001	0.1482	
p__Proteobacteria|g__Desulfovibrio	0.1095	0.1227	0.0934	0.1656	0.0296	0.3404	0.0732	0.4606	0.0069	0.2838	0.8967	
p__Proteobacteria|c__Gammaproteobacteria	0.2351	0.0103	0.0979	0.0190	0.0257	0.3738	0.0813	0.4620	0.2901	0.0000	0.2629	Healthy
p__Proteobacteria|f__Enterobacteriaceae	0.2593	0.0000	0.0558	0.0000	0.0000	NA	NA	NA	0.3322	0.0000	0.2650	
p__Proteobacteria|g__Klebsiella	0.1875	0.0423	0.1060	0.0705	0.0490	0.2593	0.0780	0.4606	0.8134	0.0000	0.2318	
p__Proteobacteria|f__Pasteurellaceae	0.2899	0.0000	0.0633	0.0000	0.0000	NA	NA	NA	0.1868	0.0001	0.3286	
p__Proteobacteria|g__Haemophilus	0.2727	0.0000	0.0644	0.0000	0.0000	NA	NA	NA	0.1931	0.0002	0.2614	
p__Tenericutes	0.2294	0.0000	0.0575	0.0001	0.0000	NA	NA	NA	0.0531	0.0000	0.0660	Healthy
p__Tenericutes|c__Mollicutes	0.2287	0.0000	0.0576	0.0001	0.0000	NA	NA	NA	0.0609	0.0000	0.0684	Healthy
p__Verrucomicrobia	0.0534	0.2880	0.0949	0.3137	0.1149	0.0738	0.0839	0.4606	0.0011	0.7013	0.0260	Healthy

Taxon	H2r	p.H2r	se.H2r	FDR.H2r	C2	p.C2	se.C2	FDR.C2	p.sex	p.age	p.MetS	Group
p__Verrucomicrobia|f__Verrucomicrobiaceae	0.0635	0.2556	0.0961	0.2858	0.0973	0.1085	0.0825	0.4606	0.0025	0.5685	0.0354	
p__Verrucomicrobia|g__Akkermansia	0.0614	0.2629	0.0963	0.2903	0.1012	0.1002	0.0828	0.4606	0.0025	0.5742	0.0360	Healthy

H2r: total additive genetic heritability; p.H2r: significance of H2r; se.H2r: H2r standard error; FDR.H2r: FDR-corrected H2r p-value;

C2: common/shared environment; p.C2: significance of C2; se.C2: C2 standard error; FDR.C2: FDR-corrected C2 p-value;

p.sex; p.age; p.MetS: significance of each covariate;

Group: association with MetS status (obtained from MaAsLin). "Healthy"-enriched in healthy individuals, "MetS"-enriched in MetS individuals

Appendix B. Comparison of the MetS-related gut microbes identified in this study and the obesity- or type 2 diabetes-related gut microbes reported previously. The taxa enriched and depleted in the individuals with a designated disease in each column compared to the healthy individuals are denoted by red and blue, respectively. Green indicates that inconsistent results concerning the association between a taxon and a disease were reported in the referenced paper.

[†]This study

[‡]Sanz Y, Olivares M, Moya-Pérez Á, Agostoni C. (2014). Understanding the role of gut microbiome in metabolic disease risk. Pediatr Res 77:236–244

[§]Tagliabue A, Elli M. (2013). The role of gut microbiota in human obesity: Recent findings and future perspectives. Nutr Metab Cardiovasc Dis 23:160-168

Kingdom	Phylum	Class	Order	Family	Genus	Species	MetS[†]	T2D[‡]	Obesity[§]
Archaea	Euryarchaeota	Methanobacteria	Methanobacteriales	Methanobacteriaceae	*Methanobrevibacter*				
Archaea	Euryarchaeota	Methanobacteria	Methanobacteriales	Methanobacteriaceae	*Methanobrevibacter*	*M. smithii*			
Bacteria	Actinobacteria								
Bacteria	Actinobacteria	Actinobacteria							
Bacteria	Actinobacteria	Actinobacteria	Bifidobacteriales						
Bacteria	Actinobacteria	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae					
Bacteria	Actinobacteria	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	*Bifidobacterium*				
Bacteria	Actinobacteria	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	*Bifidobacterium*	*Bifidobacterium spp.*			
Bacteria	Actinobacteria	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	*Bifidobacterium*	*B.animalis*			
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	[Odoribacteraceae]					
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	[Odoribacteraceae]	*Odoribacter*				
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	[Paraprevotellaceae]					
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	[Paraprevotellaceae]	*Paraprevotella*				
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae					

Kingdom	Phylum	Class	Order	Family	Genus	Species	MetS†	T2D‡	Obesity§
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	*Parabacteroides*				
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae					
Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae					
Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	*Lactobacillus*				
Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	*Lactobacillus*	*Lactobacillus spp.*			
Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	*Lactobacillus*	*L. paracasei*			
Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	*Lactobacillus*	*L. plantarum*			
Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	*Lactobacillus*	*L. reuteri*			
Bacteria	Firmicutes	Clostridia	Clostridiales	Christensenellaceae					
Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	*Dorea*				
Bacteria	Proteobacteria	Betaproteobacteria							
Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales						
Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae					
Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	*Sutterella*				
Bacteria	Proteobacteria	Gammaproteobacteria							
Bacteria	Tenericutes								
Bacteria	Tenericutes	Mollicutes							
Bacteria	Verrucomicrobia								
Bacteria	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae	*Akkermansia*				
Bacteria	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae	*Akkermansia*	*A. muciniphila*			

Appendix C. Heritability of the airway microbiota

Taxon	H2r	p.H2r	se.H2r	FDR.H2r	C2	p.C2	se.C2	FDR.C2	p.sex	p.age	p.PY
k__Bacteria	0.324	0.000	0.093	0.002	0.000	NA	NA	NA	0.714	0.708	0.165
p__Actinobacteria	0.000	0.500	NA	0.500	0.158	0.193	0.139	0.388	0.008	0.370	0.056
p__Actinobacteria|c__Actinobacteria	0.000	0.500	NA	0.500	0.149	0.199	0.138	0.388	0.009	0.365	0.070
p__Actinobacteria|o__Actinomycetales	0.000	0.500	NA	0.500	0.139	0.216	0.138	0.388	0.010	0.376	0.068
p__Actinobacteria|f__Actinomycetaceae	0.000	0.500	NA	0.500	0.164	0.248	0.131	0.388	0.009	0.800	0.031
p__Actinobacteria|g__Actinomyces	0.000	0.500	NA	0.500	0.153	0.247	0.131	0.388	0.010	0.834	0.031
p__Actinobacteria|f__Micrococcaceae	0.355	0.013	0.153	0.052	0.131	0.151	0.132	0.360	0.023	0.427	0.167
p__Actinobacteria|g__Rothia	0.355	0.013	0.153	0.052	0.131	0.151	0.133	0.360	0.023	0.420	0.172
p__Actinobacteria|f__Coriobacteriaceae	0.024	0.476	0.400	0.500	0.144	0.340	0.354	0.477	0.051	0.410	0.044
p__Actinobacteria|g__Atopobium	0.022	0.479	0.405	0.500	0.141	0.345	0.358	0.477	0.048	0.397	0.042
p__Bacteroidetes	0.593	0.000	0.122	0.002	0.000	NA	NA	NA	0.678	0.142	0.073
p__Bacteroidetes|o__Bacteroidales	0.571	0.053	0.329	0.170	0.016	0.479	0.290	0.479	0.638	0.139	0.063
p__Bacteroidetes|f__[Paraprevotellaceae]	0.394	0.024	0.192	0.086	0.117	0.230	0.165	0.388	0.163	0.035	0.729
p__Bacteroidetes|g__[Prevotella]	0.397	0.023	0.192	0.086	0.113	0.238	0.165	0.388	0.164	0.035	0.707
p__Bacteroidetes|f__Bacteroidaceae	0.546	0.000	0.072	0.000	0.000	NA	NA	NA	0.786	0.509	0.632
p__Bacteroidetes|g__Bacteroides	0.551	0.000	0.073	0.000	0.000	NA	NA	NA	0.812	0.504	0.617
p__Bacteroidetes|f__Porphyromonadaceae	0.332	0.037	0.174	0.127	0.050	0.373	0.159	0.477	0.785	0.316	0.352
p__Bacteroidetes|g__Porphyromonas	0.363	0.000	0.087	0.000	0.000	NA	NA	NA	0.674	0.362	0.269
p__Bacteroidetes|f__Prevotellaceae	0.473	0.122	0.398	0.241	0.055	0.439	0.360	0.477	0.577	0.083	0.052
p__Bacteroidetes|g__Prevotella	0.477	0.120	0.398	0.241	0.050	0.444	0.359	0.477	0.579	0.083	0.052
p__Bacteroidetes|o__Flavobacteriales	0.158	0.230	0.214	0.419	0.193	0.146	0.189	0.360	0.684	0.881	0.123
p__Bacteroidetes|f__[Weeksellaceae]	0.344	0.132	0.284	0.253	0.233	0.148	0.230	0.360	0.653	0.186	0.064
p__Bacteroidetes|f__Flavobacteriaceae	0.110	0.305	0.217	0.500	0.222	0.125	0.197	0.360	0.735	0.533	0.141

Taxon	H2r	p.H2r	se.H2r	FDR.H2r	C2	p.C2	se.C2	FDR.C2	p.sex	p.age	p.PY
p__Bacteroidetes|g__Capnocytophaga	0.108	0.308	0.217	0.500	0.220	0.127	0.197	0.360	0.733	0.544	0.147
p__Firmicutes	0.000	0.500	NA	0.500	0.371	0.026	0.122	0.240	0.061	0.904	0.001
p__Firmicutes|c__Bacilli	0.000	0.500	NA	0.500	0.156	0.130	0.142	0.360	0.034	0.354	0.035
p__Firmicutes|o__Gemellales	0.429	0.000	0.086	0.000	0.000	NA	NA	NA	0.675	0.104	0.586
p__Firmicutes|f__Gemellaceae	0.435	0.000	0.085	0.000	0.000	NA	NA	NA	0.687	0.103	0.603
p__Firmicutes|g__Gemella	0.417	0.000	0.086	0.000	0.000	NA	NA	NA	0.646	0.140	0.527
p__Firmicutes|o__Lactobacillales	0.000	0.500	NA	0.500	0.147	0.143	0.141	0.360	0.027	0.345	0.028
p__Firmicutes|f__Carnobacteriaceae	0.044	0.447	0.325	0.500	0.302	0.137	0.274	0.360	0.079	0.033	0.019
p__Firmicutes|g__Granulicatella	0.041	0.450	0.326	0.500	0.306	0.135	0.274	0.360	0.080	0.033	0.019
p__Firmicutes|f__Streptococcaceae	0.000	0.500	NA	0.500	0.118	0.195	0.140	0.388	0.036	0.552	0.039
p__Firmicutes|g__Streptococcus	0.000	0.500	NA	0.500	0.117	0.197	0.140	0.388	0.037	0.552	0.039
p__Firmicutes|c__Clostridia	0.000	0.500	NA	0.500	0.342	0.038	0.127	0.240	0.580	0.540	0.009
p__Firmicutes|o__Clostridiales	0.000	0.500	NA	0.500	0.338	0.038	0.126	0.240	0.599	0.554	0.009
p__Firmicutes|f__[Mogibacteriaceae]	0.272	0.005	0.105	0.023	0.000	NA	NA	NA	0.234	0.301	0.785
p__Firmicutes|f__Lachnospiraceae	0.299	0.005	0.116	0.023	0.000	NA	NA	NA	0.502	0.534	0.640
p__Firmicutes|g__Oribacterium	0.170	0.055	0.109	0.170	0.000	NA	NA	NA	0.874	0.301	0.506
p__Firmicutes|f__Peptostreptococcaceae	0.317	0.000	0.096	0.003	0.000	NA	NA	NA	0.771	0.334	0.158
p__Firmicutes|f__Veillonellaceae	0.000	0.500	NA	0.500	0.359	0.025	0.123	0.240	0.618	0.487	0.006
p__Firmicutes|g__Megasphaera	0.259	0.099	0.195	0.219	0.104	0.256	0.165	0.389	0.168	0.513	0.110
p__Firmicutes|g__Selenomonas	0.119	0.250	0.175	0.445	0.106	0.227	0.147	0.388	0.652	0.863	0.739
p__Firmicutes|g__Veillonella	0.000	0.500	NA	0.500	0.379	0.016	0.118	0.240	0.489	0.518	0.003
p__Fusobacteria|o__Fusobacteriales	0.066	0.336	0.157	0.500	0.000	NA	NA	NA	0.075	0.926	0.008
p__Fusobacteria|f__Fusobacteriaceae	0.000	0.500	NA	0.500	0.204	0.135	0.147	0.360	0.075	0.553	0.025
p__Fusobacteria|g__Fusobacterium	0.000	0.500	NA	0.500	0.207	0.134	0.147	0.360	0.078	0.559	0.025

Taxon	H2r	p.H2r	se.H2r	FDR.H2r	C2	p.C2	se.C2	FDR.C2	p.sex	p.age	p.PY
p__Fusobacteria|f__Leptotrichiaceae	0.238	0.105	0.173	0.219	0.035	0.404	0.149	0.477	0.898	0.584	0.137
p__Fusobacteria|g__Leptotrichia	0.225	0.104	0.167	0.219	0.100	0.234	0.145	0.388	0.426	0.482	0.234
p__Proteobacteria	0.000	0.500	NA	0.500	0.561	0.055	0.101	0.303	0.501	0.152	0.002
p__Proteobacteria|c__Betaproteobacteria	0.000	0.500	NA	0.500	0.530	0.028	0.115	0.240	0.544	0.244	0.027
p__Proteobacteria|o__Burkholderiales	0.000	0.500	NA	0.500	0.000	NA	NA	NA	0.844	0.041	0.056
p__Proteobacteria|f__Burkholderiaceae	0.000	0.500	NA	0.500	0.000	NA	NA	NA	0.953	0.132	0.065
p__Proteobacteria|f__Neisseriaceae	0.000	0.500	NA	0.500	0.531	0.030	0.115	0.240	0.552	0.227	0.029
p__Proteobacteria|g__Eikenella	0.000	0.500	NA	0.500	0.120	0.341	0.151	0.477	0.676	0.802	0.076
p__Proteobacteria|g__Neisseria	0.000	0.500	NA	0.500	0.527	0.022	0.115	0.240	0.472	0.210	0.032
p__Proteobacteria|c__Epsilonproteobacteria	0.204	0.104	0.154	0.219	0.014	0.463	0.149	0.477	0.377	0.895	0.700
p__Proteobacteria|o__Campylobacterales	0.204	0.104	0.154	0.219	0.014	0.463	0.149	0.477	0.373	0.894	0.701
p__Proteobacteria|f__Campylobacteraceae	0.206	0.100	0.153	0.219	0.014	0.462	0.149	0.477	0.378	0.895	0.705
p__Proteobacteria|g__Campylobacter	0.204	0.102	0.153	0.219	0.013	0.464	0.148	0.477	0.386	0.888	0.720
p__Proteobacteria|c__Gammaproteobacteria	0.410	0.213	0.529	0.398	0.064	0.447	0.478	0.477	0.701	0.162	0.012
p__Proteobacteria|f__Enterobacteriaceae	0.367	0.000	0.091	0.000	0.000	NA	NA	NA	0.057	0.900	0.979
p__Proteobacteria|g__Providencia	0.698	0.000	0.058	0.000	0.000	NA	NA	NA	0.627	0.263	0.398
p__Proteobacteria|f__Pasteurellaceae	0.454	0.002	0.138	0.014	0.000	NA	NA	NA	0.623	0.132	0.021
p__Proteobacteria|g__Actinobacillus	0.098	0.278	0.168	0.481	0.000	NA	NA	NA	0.904	0.390	0.070
p__Proteobacteria|g__Aggregatibacter	0.254	0.003	0.094	0.016	0.000	NA	NA	NA	0.013	0.065	0.925
p__Proteobacteria|g__Haemophilus	0.425	0.007	0.151	0.032	0.000	NA	NA	NA	0.866	0.221	0.027
p__Spirochaetes	0.222	0.092	0.161	0.219	0.010	0.467	0.124	0.477	0.065	0.613	0.749
p__Spirochaetes|c__Spirochaetes	0.222	0.092	0.161	0.219	0.010	0.466	0.124	0.477	0.063	0.591	0.757
p__Spirochaetes|f__Spirochaetaceae	0.222	0.092	0.161	0.219	0.010	0.466	0.124	0.477	0.063	0.591	0.757
p__Spirochaetes|g__Treponema	0.222	0.092	0.161	0.219	0.010	0.466	0.124	0.477	0.063	0.591	0.757

H2r: total additive genetic heritability; p.H2r: significance of H2r; se.H2r: H2r standard error; FDR.H2r: FDR-corrected H2r p-value; C2: common/shared environment; p.C2: significance of C2; se.C2: C2 standard error; FDR.C2: FDR-corrected C2 p-value; p.sex; p.age; p.PY: significance of each covariate

국문 초록 (Abstract in Korean)

인체의 유전적 요인과 환경적 요인이 장내와 호흡기 마이크로비옴에 미치는 영향

인체 마이크로비옴은 외부 환경에 노출되어 있는 모든 신체 부위에 존재하는 박테리아, 고세균, 균류, 바이러스 등의 광범위한 미생물의 집합체이다. 최근 시퀀싱 기술과 분석 기법의 급속한 발달로 인체 마이크로비옴의 구조, 기능, 숙주와의 상호작용 등에 대한 연구가 가능하게 되었다. 이를 통해, 정상 상태의 인체 마이크로비옴이 영양소 대사, 상피세포의 증식, 면역계의 발달과 유지, 병원균으로부터 보호 등 인체의 건강 유지에 필수적인 다양한 기능을 수행한다는 것이 밝혀졌다. 한편, 마이크로비옴 구조의 불균형(dysbiosis)은 비만, 암, 다양한 신체 부위에서의 염증 질환 등과 높은 상관관계가 있는 것으로 보고되고 있다. 향후 마이크로비옴 관련 질환에 대한 적절한 치료 방법 제시를 위해서는, 건강한 인구 집단에서의 정상 마이크로비옴의 특성 파악과 마이크로비옴의 구조 변화를 유발하는 요인 규명이 선행되어야 한다. 이를 위해, 본 연구에서는 한국인 쌍둥이-가족 코호트의 장내 마이크로비옴과 호흡기 마이크로비옴에 대한 대규모 연구를 수행하였다.

첫째, 건강한 한국인 일란성 쌍둥이의 36개의 분변 시료에 대해 메타게노믹 시퀀싱을 수행하여, 건강한 한국인의 장내 마이크로비옴은 두 개의 enterotype (장유형)으로 나눠지는 것을 확인하였다. 각 enterotype은 *Bacteroides*와 *Prevotella*가 우세하게 존재하였다. 개인의 enterotype은 2년 후에도 대부분 안정적으로 유지되었으며, 대부분의 쌍둥이들은 동일한 enterotype에 속하는 것으로 나타났다. 또한, enterotype은 식이섬유, 비타민, 미네랄 섭취량 등과 연관성이 있었으며, 임상의학적 지표 중 혈중 요산 수치와도 유의한 상관관계가 있는 것으로 나타났다. 이러한 결과를 통해, 인체의 유전적인 요인과 식습관 등이 개인의 enterotype을 결정짓는 중요한 요인이라는 것을 확인하였다.

둘째, 대사증후군, 장내 마이크로비옴, 유전적 요인의 연관성 규명을 위하여, 대사증후군을 포함 (18%)하는 655명의 일란성, 이란성쌍둥이, 그들의 부모, 형제의

장내 마이크로비옴 분석을 실시하였다. 대사증후군을 갖고 있는 사람의 장내에는 *Methanobrevibacter*, *Lactobacillus*, *Sutterella*가, 건강한 사람의 장내에는 *Akkermansia*, *Odoribacter*, *Bifidobacterium*이 더 많은 비율을 차지하는 것으로 나타났다. 한편, 대사증후군 관련 장내 미생물 중에서, *Bifidobacterium*이 속해있는 Actinobacteria가 가장 높은 heritability를 보였으며, 이들은 혈중중성지방 및 대사증후군과 관련된 apolipoprotein A-V gene (*APOA5*) SNP rs651821의 minor allele을 가진 사람에게서 낮은 비율로 존재하였다. 이는 특정 유전형에 의해서 장내 마이크로비옴의 구조가 변화되고, 나아가 대사증후군의 유발로 이어질 수 있다는 가능성을 제시해주는 결과이다.

마지막으로, 건강한 일란성, 이란성쌍둥이, 그들의 부모, 형제, 총 257명의 객담 시료로부터 16S rRNA gene sequencing을 실시하여, 유전적 요인과 생활습관 요인이 정상 호흡기 마이크로비옴에 미치는 영향을 연구하였다. 유전적인 요인은 *Providencia*와 *Bacteroides*를 포함하는 특정 taxa들에 영향을 끼치며, 생활습관 요인 중에서, 흡연이 호흡기 마이크로비옴 구조에 가장 큰 영향을 끼치는 것으로 확인되었다. 특히, pack-year 변수는 호흡기 염증과 관련이 있는 *Veillonella*와 유의한 양의 상관관계, *Haemophilus*와는 음의 상관관계를 보였다. 네트워크 분석에서는 흡연에 대하여 같은 경향성을 갖는 taxa들끼리 cluster를 형성하며, 유전적인 요인에 의해서만 영향을 받는 taxa들은 또 다른 cluster를 형성하는 것으로 나타났다. 이러한 결과는 호흡기 마이크로비옴의 구조는 유전적인 요인과 흡연 사이의 복잡한 상호작용에 의해 결정됨을 제시한다.

결론적으로, 본 연구는 인체 마이크로비옴의 구조는 식습관 및 생활습관 등의 인체 외적인 환경 요소뿐만 아니라 유전적 요인에 의해서도 영향을 받으며, 이러한 요인들에 의해 야기된 특정 미생물의 조성 변화는 대사증후군 또는 호흡기 염증 등과 연관성이 있을 수 있다는 것을 보여주었다. 이러한 결과는 마이크로비옴 관련 질환의 예방 및 치료에 있어서, 개인 맞춤형 접근법의 필요성을 제시해 준다.

표제어: 장내 마이크로비옴, 호흡기 마이크로비옴, 장 유형, 대사증후군, 유전적 요인

학번: 2011-30706