Role of an unclassified Lachnospiraceae in the pathogenesis of type 2 diabetes: a longitudinal study of the urine microbiome and metabolites

Kangjin Kim1, Sanghun Lee2, Sang-Chul Park3, Nam-Eun Kim1, Chol Shin4,5, Seung Ku Lee5, Youngae Jung6, Dankyu Yoon7, Hyeonjeong Kim8, Sanghyun Kim8, Geum-Sook Hwang6,7,8 and Sungho Won1,9,10

© The Author(s) 2022

Recent investigations have revealed the human microbiome plays an essential role in the occurrence of type 2 diabetes (T2D). However, despite the importance of understanding the involvement of the microbiota throughout the body in T2D, most studies have focused specifically on the intestinal microbiota. Extracellular vesicles (EVs) have been recently found to provide important evidence regarding the mechanisms of T2D pathogenesis, as they act as key messengers between intestinal microorganisms and the host. Herein, we explored microorganisms potentially associated with T2D by tracking changes in microbiota-derived EVs from patient urine samples collected three times over four years. Mendelian randomization analysis was conducted to evaluate the causal relationships among microbial organisms, metabolites, and clinical measurements to provide a comprehensive view of how microbiota can influence T2D. We also analyzed EV-derived metagenomic (N = 393), clinical (N = 5032), genomic (N = 8842), and metabolite (N = 574) data from a prospective longitudinal Korean community-based cohort. Our data revealed that GU174097_g, an unclassified Lachnospiraceae, was associated with T2D (β = −189.13; p = 0.00006), and it was associated with the ketone bodies acetoacetate and 3-hydroxybutyrate (r = −0.0938 and −0.0829, respectively; p = 0.0022 and 0.0069, respectively). Furthermore, a causal relationship was identified between acetoacetate and HbA1c levels (β = 0.0002; p = 0.154). GU174097_g reduced ketone body levels, thus decreasing HbA1c levels and the risk of T2D. Taken together, our findings indicate that GU174097_g may lower the risk of T2D by reducing ketone body levels.

 Experimental & Molecular Medicine (2022) 54:1125–1132; https://doi.org/10.1038/s12276-022-00816-x

INTRODUCTION

Recent studies have revealed that the intestinal microbiota plays essential roles in host energy homeostasis, body adiposity, blood sugar control, insulin sensitivity, hormone secretion, and the pathogenesis of metabolic diseases, such as type 2 diabetes (T2D) and obesity1–3. However, most of these studies analyzed stool samples and therefore obtained limited information relative to insights from direct sampling of the intestinal mucosa, which is not possible in most cases. In addition, the composition of microbial communities in stool samples is greatly affected by the specific compartment in which they reside, such as the mucous membrane. Microbial communities also differ based on their source, ranging from the intestines, skin, and airways, which are frequently studied, to urine and blood, which are generally sterile environments. Therefore, it is important not only to understand the role of the intestinal microbiota but also to consider the function and combined contribution of the all microbiota throughout the body.

Extracellular vesicles (EVs) have been recently suggested to function as the main messengers between intestinal microorganisms and the host. EVs travel long distances within and between body tissues6 and have been used as biomarkers of atopic dermatitis, alcoholic hepatitis, and asthma7–10. Microbiota-derived EVs can enter the circulatory system through the intestinal barrier. They are suspected to play a key role in the development of insulin resistance, potentially providing important clues into the pathogenesis of T2D. For example, EVs derived from Pseudomonas panacis are present in the stool samples of high-fat diet-fed mice. They can infiltrate the gut barrier and block the insulin pathway in skeletal muscle and adipose tissue, inducing the development of insulin resistance and glucose intolerance11. However, microbiota-derived EVs are highly variable, as they are modulated by different factors, such as age and sex. Therefore, caution should be exercised when inferring causal relationships based on the statistical analysis of microbiota-derived data. Furthermore,
longitudinal microbiota studies may allow for stronger inferences than cross-sectional studies and may allow for the detection of microorganisms related to the progression of T2D in healthy subjects. However, existing studies have been predominantly cross-sectional in nature and are based on correlation analyses. As a result, these studies are unable to comprehensively provide an understanding of the exact roles of the intestinal microbiota and EVs in metabolic disease development.

Therefore, in the present study, we investigated the prospective Korean Association REsource project (KARE) cohort. By tracking changes in microbiota-derived EVs in urine samples from Korean adults collected three times over four years, we explored the potential associations between microorganisms and T2D progression. Furthermore, using genomic and metabolite data from the KARE cohort, we conducted a multimomics analysis to investigate the specific role of microorganisms potentially involved in T2D pathogenesis. We expect our findings to provide information regarding how microbes, the substances they produce, and their byproducts interact with the human body and affect metabolic disease development. In addition, we evaluated causal relationships among microbial organisms, ketone bodies, and clinical measurements, with the aim of further elucidating the relationship between T2D and the microbiota.

MATERIALS AND METHODS

Cohort and study design

The KARE cohort is a prospective study cohort involving subjects from the rural community of Ansong and the urban community of Ansan in South Korea. The KARE project began in 2001 as part of the Korean Genome Epidemiology Study. We used data from urine samples taken from Korean adults collected three times over four years, the potential associations between microorganisms and T2D progression. Furthermore, we conducted a multimomics analysis to investigate the specific role of microorganisms potentially involved in T2D pathogenesis. We expect our findings to provide information regarding how microbes, the substances they produce, and their byproducts interact with the human body and affect metabolic disease development. In addition, we evaluated causal relationships among microbial organisms, ketone bodies, and clinical measurements, with the aim of further elucidating the relationship between T2D and the microbiota.

Association analysis with T2D

- Metagenomic and clinical data: relative abundance of genera, T2D status and diabetes risk indicators.
- Functional profile and clinical data: abundances of functional profiles and T2D status.

Network analysis of G174907_g

- Metagenomic and clinical data: relative abundance of G174907_g and T2D risk indicators.
- Clinical and metabolites data: T2D risk indicators and abundance of acetocetate and 3-hydroxybutyrate.
- Metagenomic and metabolites data: relative abundance of G174907_g and abundance of acetocetate and 3-hydroxybutyrate.

GWAS with population 1

- Metagenomic and genomic data

GWAS with population 2

- Metabolites and genomic data

GWAS with population 3

- Clinical and genomic data

Operational definition of T2D and related phenotypes

Study participants were categorized into control individuals, T2D-at-risk patients, and T2D patients. T2D and T2D-at-risk patients were diagnosed on the basis of the American Diabetes Association criteria, which are provided in Supplementary Table 1. T2D status was then stratified into T2D-at-risk/T2D (0 for healthy; 1 for T2D-at-risk and T2D) and binary_T2D (0 for healthy and T2D-at-risk; 1 for T2D). In addition, we considered other T2D-related indicators, such as BMI, HbA1c levels, fasting glucose and insulin levels, 60- and 120-min plasma glucose levels, and insulin levels after a 75 g oral glucose tolerance test in our analysis. Age, the levels of total cholesterol, high-density lipoprotein (HDL) cholesterol, triglycerides, kidney- and liver-related disease indicators (blood urea nitrogen (BUN), creatinine, aspartate aminotransferase (AST), and alanine aminotransferase (ALT) C-reactive protein (CRP), white blood cell (WBC) count, red blood cell (RBC) count, hemoglobin, hematocrit, and platelet count) were also collected. The homeostatic model assessment for insulin resistance (HOMA-IR) was calculated using fasting glucose and fasting insulin levels. Descriptive statistics for all variables were generated using Rex software (RexSoft Inc., Seoul, Korea) (Supplementary Table 2).

EV isolation and DNA extraction

For EV isolation, urine samples were subjected to differential centrifugation at 10,000 x g and 4 °C for 10 min using a microcentrifuge (Labogene 1730R; Bio-Medical Science, Seoul, Korea). To remove bacteria, foreign particles, and waste, the supernatant was filtered through a 0.22-µm filter (Inchpor2 Syringe Filter; Inchemtec, Seoul, Korea). The isolated EVs were boiled at 100 °C for 40 min and centrifuged at 18,214 x g and 4 °C for 30 min to eliminate floating particles and impurities. The supernatant was collected and subjected to DNA extraction using a PowerSoil® DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA) according to the manufacturer’s protocol. DNA was quantified using the QIAxpert system (Qiagen, Hilden, Germany).

16S rRNA sequence data processing

Paired-end sequencing of the V3-V4 region of the bacterial 16S rRNA gene was conducted at MD Health care (Seoul, Korea) with the MiSeq Reagent

SPRINGER NATURE

Experimental & Molecular Medicine (2022) 54:1125 – 1132
Analyses from 393 subjects, including 70 genera, were used for subsequent network analysis. PERMANOVA can be applied to the cross-sectional data, and thus, the beta-diversity, which, in our study, was calculated using the Vegan package in R (v2.5.6). Taxonomy-based ring charts were retained for analysis.

Network analysis of a T2D-related taxon based on multomics data

To assess overall associations using repeatedly measured multomics data, we first modeled an LMM using log-transformed diabetes risk indicators as response variables and age in phase 1 as well as sex as explanatory variables with a compound symmetry structure for its covariance. We modeled an LMM with a T2D-related taxon as the response variable with the same covariates and covariance structure. For each combination of diabetes risk indicators and a T2D-related taxon, two different sets of residuals were obtained, and Spearman correlations between the residuals were calculated. Similarly, the association between a chosen microbial marker and the levels of ketone bodies was analyzed. Network analysis was conducted to calculate simple correlations among diabetes risk indicators, a chosen taxon, and ketone bodies. Edge width was calculated as \(-\log_{10}\) of the \(p\) value. The network was visualized using the R package visNetwork (v2.2.0).

Genotypic, imputation, and quality control

Quality control and genotype imputation were performed according to the standard quality control and imputation protocols for the genotypes of 8842 KARE cohort participants. After quality control, 8216 subjects with 17,716,215 single-nucleotide polymorphisms (SNPs) were included in the analysis. In total, the data of 351 subjects with a read count <3,000 and nonmissing T2D status for all phases were used for a genome-wide association study (GWAS) of metagenomic data. A total of 574 subjects who had no missing metabolite levels and T2D status for all three phases were selected for a GWAS of metabolite levels. Among the subjects not included in the metabolite or metagenome GWAS, 3542 subjects had KARE phenotypes for the three phases and were thus included in a GWAS of KARE phenotypes. We excluded subjects in the metabolite or metagenome GWAS for the purposes of a two-sample Mendelian randomization (MR) study. Details are provided in Supplementary Fig. 1, and all the associated SNPs from each GWAS are listed in Supplementary Table 3.

MR analysis

MR uses genetic variants that are not associated with conventional confounders of observational studies and is therefore considered analysis that is not randomized controlled trials. Randomly selected alleles with a compound symmetry structure for each time point was incorporated to adjust the similarity of T2D status for the same subject at different time points, and the sandwich estimator was used to find a robust estimate against the misspecified covariance matrix. To accommodate the multiple testing problem, \(p\) values were adjusted for the false discovery rate (FDR) using the Benjamini–Hochberg method.

Network analysis of a T2D-related taxon based on multomics data

To assess overall associations using repeatedly measured multomics data, we first modeled an LMM using log-transformed diabetes risk indicators as response variables and age in phase 1 as well as sex as explanatory variables with a compound symmetry structure for its covariance. We modeled an LMM with a T2D-related taxon as the response variable with the same covariates and covariance structure. For each combination of diabetes risk indicators and a T2D-related taxon, two different sets of residuals were obtained, and Spearman correlations between the residuals were calculated. Similarly, the association between a chosen microbial marker and the levels of ketone bodies was analyzed. Network analysis was conducted to calculate simple correlations among diabetes risk indicators, a chosen taxon, and ketone bodies. Edge width was calculated as \(-\log_{10}\) of the \(p\) value. The network was visualized using the R package visNetwork (v2.2.0).

Genotypic, imputation, and quality control

Quality control and genotype imputation were performed according to the standard quality control and imputation protocols for the genotypes of 8842 KARE cohort participants. After quality control, 8216 subjects with 17,716,215 single-nucleotide polymorphisms (SNPs) were included in the analysis. In total, the data of 351 subjects with a read count <3,000 and nonmissing T2D status for all phases were used for a genome-wide association study (GWAS) of metagenomic data. A total of 574 subjects who had no missing metabolite levels and T2D status for all three phases were selected for a GWAS of metabolite levels. Among the subjects not included in the metabolite or metagenome GWAS, 3542 subjects had KARE phenotypes for the three phases and were thus included in a GWAS of KARE phenotypes. We excluded subjects in the metabolite or metagenome GWAS for the purposes of a two-sample Mendelian randomization (MR) study. Details are provided in Supplementary Fig. 1, and all the associated SNPs from each GWAS are listed in Supplementary Table 3.

MR analysis

MR uses genetic variants that are not associated with conventional confounders of observational studies and is therefore considered analysis that is not randomized controlled trials. Randomly selected alleles with a compound symmetry structure for each time point was incorporated to adjust the similarity of T2D status for the same subject at different time points, and the sandwich estimator was used to find a robust estimate against the misspecified covariance matrix. To accommodate the multiple testing problem, \(p\) values were adjusted for the false discovery rate (FDR) using the Benjamini–Hochberg method.

Network analysis of a T2D-related taxon based on multomics data

To assess overall associations using repeatedly measured multomics data, we first modeled an LMM using log-transformed diabetes risk indicators as response variables and age in phase 1 as well as sex as explanatory variables with a compound symmetry structure for its covariance. We modeled an LMM with a T2D-related taxon as the response variable with the same covariates and covariance structure. For each combination of diabetes risk indicators and a T2D-related taxon, two different sets of residuals were obtained, and Spearman correlations between the residuals were calculated. Similarly, the association between a chosen microbial marker and the levels of ketone bodies was analyzed. Network analysis was conducted to calculate simple correlations among diabetes risk indicators, a chosen taxon, and ketone bodies. Edge width was calculated as \(-\log_{10}\) of the \(p\) value. The network was visualized using the R package visNetwork (v2.2.0).

Genotypic, imputation, and quality control

Quality control and genotype imputation were performed according to the standard quality control and imputation protocols for the genotypes of 8842 KARE cohort participants. After quality control, 8216 subjects with 17,716,215 single-nucleotide polymorphisms (SNPs) were included in the analysis. In total, the data of 351 subjects with a read count <3,000 and nonmissing T2D status for all phases were used for a genome-wide association study (GWAS) of metagenomic data. A total of 574 subjects who had no missing metabolite levels and T2D status for all three phases were selected for a GWAS of metabolite levels. Among the subjects not included in the metabolite or metagenome GWAS, 3542 subjects had KARE phenotypes for the three phases and were thus included in a GWAS of KARE phenotypes. We excluded subjects in the metabolite or metagenome GWAS for the purposes of a two-sample Mendelian randomization (MR) study. Details are provided in Supplementary Fig. 1, and all the associated SNPs from each GWAS are listed in Supplementary Table 3.

MR analysis

MR uses genetic variants that are not associated with conventional confounders of observational studies and is therefore considered analysis that is not randomized controlled trials. Randomly selected alleles with a compound symmetry structure for each time point was incorporated to adjust the similarity of T2D status for the same subject at different time points, and the sandwich estimator was used to find a robust estimate against the misspecified covariance matrix. To accommodate the multiple testing problem, \(p\) values were adjusted for the false discovery rate (FDR) using the Benjamini–Hochberg method.

Network analysis of a T2D-related taxon based on multomics data

To assess overall associations using repeatedly measured multomics data, we first modeled an LMM using log-transformed diabetes risk indicators as response variables and age in phase 1 as well as sex as explanatory variables with a compound symmetry structure for its covariance. We modeled an LMM with a T2D-related taxon as the response variable with the same covariates and covariance structure. For each combination of diabetes risk indicators and a T2D-related taxon, two different sets of residuals were obtained, and Spearman correlations between the residuals were calculated. Similarly, the association between a chosen microbial marker and the levels of ketone bodies was analyzed. Network analysis was conducted to calculate simple correlations among diabetes risk indicators, a chosen taxon, and ketone bodies. Edge width was calculated as \(-\log_{10}\) of the \(p\) value. The network was visualized using the R package visNetwork (v2.2.0).
including IVW, MR-egger, MR-egger with SIMEX correction, median-weighted method, and MR-PRESSO, and selected the recommended MR method based on the violations of MR assumptions36.

RESULTS

Longitudinal changes in the urine microbial composition over four years

The alpha-diversity of the urine microbiome decreased during the follow-up period, which may have been an effect of aging (Supplementary Fig. 2). A nonmetric multidimensional scaling plot based on beta diversity also revealed a gradual change in microbiota composition with age (Supplementary Fig. 3). The overall microbiome composition at the phylum and genus levels is presented in Fig. 2 and Supplementary Fig. 4, respectively. Verrucomicrobia, Bacteroidetes, and Firmicutes were the predominant phyla, whereas Akkermansia and Bacteroides were the predominant genera.

T2D and other clinical traits explained by microbial variance

We investigated the associations between various clinical phenotypes and microbial compositions using PERMANOVA (Supplementary Fig. 5). HbA1c, WBC, hematocrit, binary_T2D, and age in phase 1 significantly explained changes in microbial composition during the follow-up period ($p = 0.0061, 0.0107, 0.0110, 0.0409, \text{ and } 0.0290$, respectively; FDR-adjusted $p = 0.1027, 0.1027, 0.1027, 0.2290, \text{ and } 0.2030$, respectively). HbA1c and binary_T2D partially explained the variance in microbial changes over the 4 years, indicating that the longitudinal change in microbiome composition may be more closely associated with T2D-related phenotypes than with other clinical traits.

Taxa and functional profiles associated with T2D and diabetes risk indicators

In an association analysis of 70 genera with binary_T2D and T2D-at-risk/T2D phenotypes, GU174097_g, an unclassified Lachnospiraceae, was found to exhibit a significant association with these phenotypes and was more abundant in healthy subjects than in diabetic or prediabetic patients (Table 1). We divided the samples into four groups. The Healthy in Phases 1-3 group included subjects who were healthy in phases 1, 2, and 3. The T2D in Phases 1-3 group consisted of subjects who had T2D in phases 1, 2, and 3. The Healthy to T2D-at-risk/T2D group included subjects who were healthy in phase 1 and became T2D patients or T2D-at-risk in phase 3. The T2D-at-risk/T2D to Healthy group included subjects who were T2D-at-risk/T2D in phase 1 and healthy in phase 3. The relative abundance of GU174097_g in subjects who were healthy at baseline but changed to the Healthy group exhibited no tendency to decrease ($p = 0.19$) (Fig. 3). Supplementary Fig. 6 shows the profiles of GU174097_g for randomly selected subjects. The relative abundance of GU174097_g in subjects who were healthy at baseline but changed to the T2D-at-risk/T2D group at phase 2 or 3 decreased with the development of T2D ($p = 0.0001$). Conversely, its relative abundance in the T2D-at-risk/T2D to Healthy group increased over time, and this association was not simply based on diabetic or nondiabetic status.

To investigate the T2D-associated microbial functional profiles, 238 functional profiles were evaluated. The significant associations at an FDR-adjusted significance of 0.1 are presented in Supplementary Table 4. The T2D-at-risk/T2D phenotype was related to...
the cationic antimicrobial peptide. Furthermore, the biosynthesis of fatty acids, coenzyme A (CoA), and secondary metabolites as well as oxidative phosphorylation were significantly associated with the Binary_T2D phenotype at an FDR-adjusted significance of 0.1.

Next, we investigated the associations between the log-transformed diabetes risk indicators and genera, and significant associations at an FDR-adjusted significance of 0.1 were identified. Twelve, four, and 20 genera were significantly associated with HbA1c, glucose, and insulin levels, respectively. In particular, *Hafnia* was associated with HbA1c and 60- and 120-min insulin levels, whereas *AB185816_g* and *Akkermansia* were associated with HbA1c, fasting glucose, and 60-min insulin levels (Supplementary Table 5).

**Table 1.** Analysis of the associations between type 2 diabetes (T2D) and bacterial genera.

| Phenotype          | Genus             | Estimate | Std Err | DF  | p value     | FDR         |
|--------------------|-------------------|----------|---------|-----|-------------|-------------|
| T2D-at-risk/T2D    | GU174097_g        | −189.13  | 46.63   | 735 | 0.00006     | 0.00393     |
| Binary_T2D         | JN713389_g        | −13.07   | 5.31    | 735 | 0.01411     | 0.38195     |
| Binary_T2D         | Akkermansia       | −3.49    | 1.43    | 735 | 0.01489     | 0.38195     |
| Binary_T2D         | Dialister         | −86.44   | 37.49   | 735 | 0.02140     | 0.38195     |
| Binary_T2D         | Ruminococcus_g2   | −25.38   | 11.70   | 735 | 0.03039     | 0.38195     |
| Binary_T2D         | KE159538_g        | −48.29   | 22.95   | 735 | 0.03568     | 0.38195     |
| Binary_T2D         | Bifidobacterium   | 6.71     | 3.21    | 735 | 0.03669     | 0.38195     |
| Binary_T2D         | Eubacterium_g8    | −71.10   | 34.46   | 735 | 0.03944     | 0.38195     |
| Binary_T2D         | Megamonas         | −65.20   | 32.53   | 735 | 0.04538     | 0.38195     |
| Binary_T2D         | Pseudomonas       | 7.74     | 3.91    | 735 | 0.04842     | 0.38195     |

Associations between genera and T2D-at-risk/T2D and Binary_T2D were tested, and significant associations at a significance level of 0.05 are summarized.

**Fig. 3** Relative proportions of GU174097_g in different type 2 diabetes (T2D) groups. The mean relative proportions of GU174097_g are provided for the Healthy in Phases 1-3, T2D in Phases 1-3, Healthy to T2D-at-risk/T2D, and T2D-at-risk/T2D to Healthy groups and are compared according to the T2D status. The Healthy in Phases 1-3 group included subjects who were healthy in phases 1, 2, and 3. The T2D in Phases 1-3 group included subjects who had T2D in phases 1, 2, and 3. The Healthy to T2D-at-risk/T2D group included subjects who were healthy in phase 1 and became T2D patients or T2D-at-risk in phase 3. The T2D-at-risk/T2D to Healthy group included subjects who were T2D-at-risk/T2D in phase 1 and healthy in phase 3. The p values were calculated based on simple linear regression using the log relative proportion of GU174097_g as a response variable and phase variable. The relative proportion of GU174097_g was log-transformed after adding one to avoid zero values. The phase variable is coded by 1, 2, and 3 for phases 1, 2, and 3, respectively.

Associations between T2D-related unclassified Lachnospiraceae and diabetes risk indicators and ketone bodies

To confirm the association between GU174097_g and T2D, we performed extensive validation analysis using clinical and metabolite data. We analyzed the association between GU174097_g and diabetes risk indicators (Table 2). Among all glucose- and insulin-related variables, GU174097_g was significantly and positively associated with the 60-min insulin level.

Thereafter, we analyzed the potential associations between ketone bodies and the T2D-related taxon, since ketone bodies have been suggested as markers of disrupted glucose metabolism in prediabetic patients37. The ketone bodies 3-hydroxybutyrate and acetoacetate exhibited significant negative correlations with...
One-sample MR did not reveal any significant association between diabetes risk indicators and ketone bodies.

**Causal relationship between the T2D-related taxon and ketone bodies and the diabetes risk indicators**

One-sample MR did not reveal any significant causal relationship between **GU174097_g** and ketone bodies and vice versa (Supplementary Table 6). To verify whether a causal relationship existed between the abundance of **GU174097_g** or the levels of ketone bodies and diabetes risk indicators, two-sample MR analysis was performed. Extensive assumption checks were conducted to enhance the validity of the two-sample MR analysis (Supplementary Table 7). No weak instrument bias was observed assuming that the IVW method—robust approach used in cases of InSIDE assumption violation—has to be considered with each recommended method. Therefore, MR-Egger (SIMEX) was used to estimate the causal effect of 3-hydroxybutyrate on 60-min insulin as well as that of acetocacetate on HbA1c levels. The IVW method was used to estimate all other causal effects. To determine the causal effect of 3-hydroxybutyrate on 60-min insulin, rs2259835 was detected as an outlier via MR-PRESSO at a significance level of 0.05 (Supplementary Table 8). Thus, rs2259835 had to be removed to prevent potential horizontal pleiotropy. The result of MR-PRESSO is shown in Supplementary Table 9 and shows the estimates without outliers. The effect of acetocacetate on the HbA1c level was the only significant effect at an FDR-adjusted significance of 0.05, indicating that acetocacetate increases HbA1c levels (Supplementary Table 9). The results obtained using the weighted median method corroborated this significant association ($p = 0.0475$).

**DISCUSSION**

Recent microbiome studies have shown that T2D is associated with gut dysbiosis that can result in altered intestinal barrier function and a dysregulation of host metabolic and signaling pathways. Intestinal bacteria can promote insulin resistance by triggering inflammation via polysaccharides, which are components of the gram-negative bacterial cell wall. Furthermore, microbiota-derived EVs are expected to affect insulin resistance and provide a more in-depth understanding of T2D pathogenesis. Various bacterial metabolites, such as short-chain fatty acids (SCFAs), can modulate the function of various signaling pathways implicated in human health and can protect against insulin resistance.

The human microbiota is highly variable, and this variability is determined by various external factors, such as diet, exercise, mobility, medication, and microbial cooccurrence patterns. Many of these external factors also determine the risk of metabolic disease and are age-related; that is, the intestinal microbiota and host phenotype are substantially altered with aging. Furthermore, the effect of the intestinal microbiota on the host phenotype is also dependent on the age of the host. The estimation of within-subject covariate effects represents a robust approach against between-subject confounders, and longitudinally measured microbiome data enable characterization of the effects of the microbiota on host disease risk. As most existing studies have been cross-sectional in nature, the validity and interpretation of their results are limited. In turn, longitudinal studies are needed to comprehensively investigate the association between the human microbiome and diseases, including T2D.

Our longitudinal study revealed that a low abundance of **GU174097_g** is a risk factor for T2D development. **GU174097_g** has not been cultured to date. Multimics data, including host genomic data, T2D-related metabolites, clinical information, and predicted functional metagenomic profiles, were utilized to extensively validate our results via causality analysis. **GU174097_g** is a member of the family Lachnospiraceae and an association between *Lachnospiraceae* and T2D risk has been reported in several previous studies. SCFA pathways, including the propanediol and acrylate signaling pathways, play important roles in mediating the effects of *Lachnospiraceae* on T2D. Additionally, SCFA-producing bacteria affect epigenetic regulation in T2D patients and reduce the risk of developing T2D. We found that **GU174097_g** is positively correlated with the 60-min insulin level, and in turn, it is negatively correlated with HbA1c levels. This indicates that **GU174097_g** reduces HbA1c levels and, thus, the risk of developing T2D by stimulating insulin secretion. Next, we aimed to elucidate how **GU174097_g** affects T2D through the regulation of 60-min insulin and HbA1c. Multiple mechanisms may underlie these associations, including the effects...
of various microbiota-derived metabolites, including SCFAs, as previously suggested. In addition, ketone bodies have been reported not only as indicators of diabetic hyperglycemia but also as markers of disturbed glucose metabolism in the prediabetic state. Furthermore, fatty acid metabolism, CoA synthesis, and oxidative phosphorylation, all of which are involved in ketogenesis or ketolysis, have been associated with T2D. In our study, the ketone bodies 3-hydroxybutyrate and acetoacetate were negatively correlated with GU174097_g but positively correlated with the 60- and 120-min glucose levels. MR analysis was employed to investigate the effects of GU174097_g and ketone bodies on diabetes risk indicators. Although no causal relationship was observed between GU174097_g and ketone bodies or other clinical variables, acetoacetate was found to be causally related to an increased HbA1c level. HbA1c level is a major biomarker of T2D and explains the microbial beta-diversity. Furthermore, GU174097_g was negatively correlated with acetoacetate. Therefore, our study not only confirmed the importance of ketone bodies in T2D pathogenesis but also suggests an underlying mechanism for the association between GU174097_g and T2D development.

Previous studies have reported that gut microbe–derived EVs can infiltrate the circulatory system through the gut barrier. Furthermore, microbe-derived EVs in urine can reflect the lung and gut microbiota of children with asthma. Interestingly, T2D increases the co-occurrence of the same OTUs within the gut microbiome and microbe-derived EVs in urine samples, which indicates that these EVs may reflect the gut microbiota composition. Coprococcus, a member of the Lachnospiraceae family, is one of the major butyrate-producing bacteria. It is known to utilize metabolic intermediates essential for the synthesis of ketone bodies, such as acetoacetyl-CoA, 3-hydroxybutyryl-CoA, and crotonyl-CoA. As energy sources to produce the SCFA butyrate, SCFAs are considered beneficial for health and are considered to protect against T2D. Thus, we hypothesize that GU174097_g consumes acetoacetate to produce SCFAs. These SCFAs can promote insulin secretion and decrease HbA1c levels, leading to a decreased risk of T2D.

Our study had several limitations. First, as it was based on the metagenomic profiles of EVs, the microbial compositions observed can differ from, and need to be further compared with, those of the intestinal microbiota. Second, as the genus-level taxonomy of GU174097_g is unknown, ecological and biological information on this species is limited. Third, published summary statistics of microbial GWAS are limited, and the sample size in the current microbial GWAS was small. Therefore, the number of SNPs reported not only as indicators of diabetic hyperglycemia but also as predictors of T2D were highly correlated and interacted with each other. Ketone bodies and diabetes risk indicators were further identified and characterized. Fourth, even though extensive methods were used to validate assumptions in our MR analysis and enhance the validity of causal analysis, the MR results were not easy to interpret. Ketone bodies and diabetes risk indicators were highly correlated and interacted with each other. Additional in vivo and in vitro experiments may clarify the associations identified herein.

Our study revealed that GU174097_g, an unclassified Lachnospiraceae, is associated with T2D and ketone bodies. Furthermore, we found a potential causal relationship between ketone body acetoacetate and HbA1c levels. Our findings indicate that GU174097_g may lower the risk of developing T2D via the reduction in ketone body levels. Although the mechanisms by which GU174097_g and ketone bodies affect T2D development have not been elucidated, further large-scale longitudinal studies as well as in vivo and in vitro experiments could contribute to unraveling these mechanisms.
28. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 26, 32–46 (2001).
29. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B Met. 57, 289–300 (1995).
30. Gim, J. et al. A between ethnicities comparison of chronic obstructive pulmonary disease genetic risk. Front. Genet. 11, 329 (2020).
31. Greco M, F. D., Minelli, C., Sheehan, N. A. & Thompson, J. R. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. Stat. Med. 34, 2926–2940 (2015).
32. Richard, M. A. et al. Altered mechanisms of genital development identified through integration of DNA methylation and genomic measures in hypospadias. Sci. Rep. 10, 1–12 (2020).
33. Bell, J. A. et al. Influence of puberty timing on adiposity and cardiometabolic traits: a Mendelian randomisation study. PLoS Med. 15, e1002641 (2018).
34. Hausman, J. A. Specification tests in econometrics. Econometrica 1251–1271 (1978).
35. Jin, H., Lee, S. & Won, S. Causal evaluation of laboratory markers in Type 2 Diabetes on cancer and vascular diseases using various mendelian randomization tools. Front. Genet. 11, S97420 (2020).
36. Mahendran, Y. et al. Association of ketone body levels with hyperglycemia and type 2 diabetes in 9,398 Finnish men. Diabetes 62, 3618–3626 (2013).
37. Bowden, J. Misconceptions on the use of MR-Egger regression and the evaluation of the InSIDE assumption. Int. J. Epidemiol. 46, 2097–2099 (2017).
38. Wen, L. & Duffy, A. Factors influencing the gut microbiota, inflammation, and type 2 diabetes. J. Nutr. 147, 14685–14755 (2017).
39. Cani, P. D., Osto, M., Geurts, L. & Everard, A. Involvement of gut microbiota in the development of low-grade inflammation and type 2 diabetes associated with obesity. Gut microbes 3, 279–288 (2012).
40. Upadhyaya, S. & Banerjee, G. Type 2 diabetes and gut microbiome: at the intersection of known and unknown. Gut microbes 6, 85–92 (2015).
41. Sharma, S. & Tripathi, P. Gut microbiome and type 2 diabetes: where we are and where to go? J. Nutr. Biochem. 63, 101–108 (2019).
42. Vallianou, N. G., Stratigou, T. & Tsagarakis, S. Microbiome and diabetes: where are we now? Diabetes Res. Clin. Pract. 146, 111–118 (2018).
43. Puddu, A., Sanguineti, R., Montecucco, F. & Viviani, G. L. Evidence for the gut microbiota short-chain fatty acids as key pathophysiological molecules improving diabetes. Mediat. Inflamm. 2014, 162021 (2014).
44. O’Toole, P. W. & Jeffery, I. B. Gut microbiota and aging. Science 350, 1214–1215 (2015).
45. Vacca, M. et al. The controversial role of human gut lachnospiraceae. Microorganisms 8, 573 (2020).
46. Ibrahim, K. S. et al. Characterisation of gut microbiota of obesity and type 2 diabetes in a rodent model. Biosci. Microbes Food Health 40, 2019–2031 (2020).
47. Sharon, G. et al. Specialized metabolites from the microbiome in health and disease. Cell Metab. 20, 719–730 (2014).
48. Kohlmeier, M. Nutrient Metabolism: Structures, Functions, and Genetics, in Food Science and Technology International Series 1st edn, 147–148 (Academic Press, London, 2003).
49. Kim, M. R. et al. S taphylococcus aureus-derived extracellular vesicles induce neutrophilic pulmonary inflammation via both T h1 and T h17 cell responses. Allergy 67, 1271–1281 (2012).
50. Lee, Y. S., Kim, J. H. & Lim, D. H. Urine microbe-derived extracellular vesicles in children with asthma. Allergy Asthma Immunol. Res. 13, 75 (2021).
51. Nah, G. et al. Type-2 diabetics reduces spatial variation of microbiome based on extracellular vesicles from gut microbes across human body. Sci. Rep. 9, 1–10 (2019).
52. Sasaki, K., Sasaki, D., Hannyta, A., Tsujioka, J. & Kondo, A. In vitro human colonic microbiota utilise D-β-hydroxybutyrate to increase butyrogenesis. Sci. Rep. 10, 1–8 (2020).

ACKNOWLEDGEMENTS
This work was supported by the Industrial Core Technology Development Program (20000134) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea). This study was conducted with bioresources from National Biobank of Korea, the Korea Disease Control and Prevention Agency, Republic of Korea (NBK-2020-101).

COMPETING INTERESTS
The authors declare no competing interests.

ETHICS APPROVAL
The protocol used in this study was approved by the Institutional Review Board (IRB No. E1801/001-004) of Seoul National University.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s12276-022-00816-x.

Correspondence and requests for materials should be addressed to Geum-Sook Hwang or Sungho Won.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022