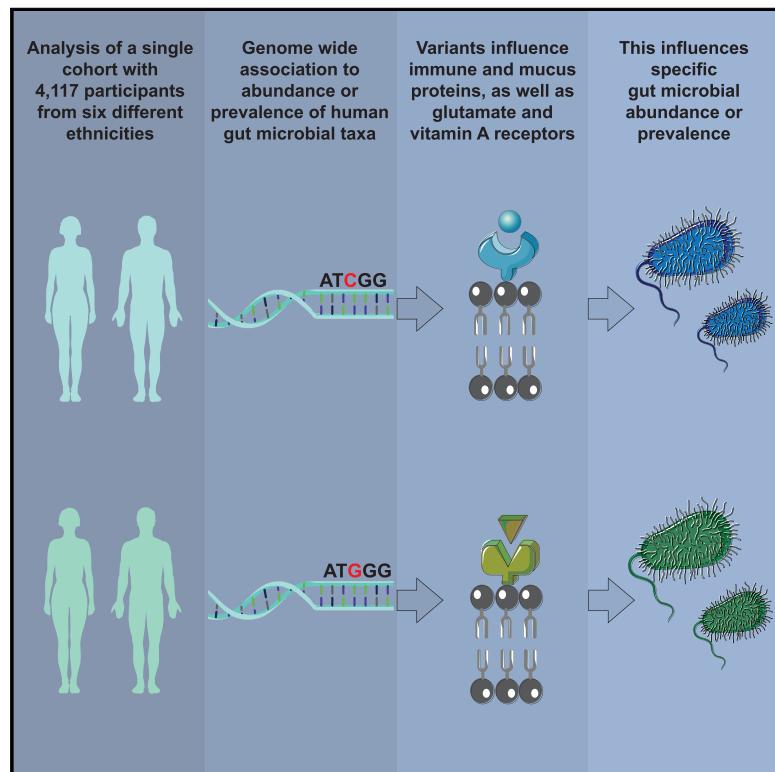


Clinical and Translational Report

Cell Host & Microbe

Gut microbiome associations with host genotype vary across ethnicities and potentially influence cardiometabolic traits

Graphical abstract



Authors

Ulrika Boulund, Diogo M. Bastos,
Bart Ferwerda, ..., Albert K. Groen,
Aeilko H. Zwinderman, Max Nieuwdorp

Correspondence

m.nieuwdorp@amsterdamumc.nl

In brief

The human genome associates with the gut microbiota, but this remains understudied in different ethnicities. Boulund et al. identified host-genome-to-gut-microbe associations in 4,117 subjects of six different ethnicities. These loci are related to immune functions, glutamate, and the mucus layer. Some loci exhibited contrasting associations between ethnicities.

Highlights

- Different ethnicities exhibit contrasting gut-microbe-to-genotype associations
- Microbe-associated loci are involved in immune functions or linked with glutamate
- Genetic loci in genes expressed in the gut, mucus, or brain are microbe associated
- Gut microbes may influence cardiometabolic health factors such as cholesterol and BMI



Clinical and Translational Report

Gut microbiome associations with host genotype vary across ethnicities and potentially influence cardiometabolic traits

Ulrika Boulund,¹ Diogo M. Bastos,¹ Bart Ferwerda,² Bert-Jan van den Born,^{1,3} Sara-Joan Pinto-Sietsma,^{1,2} Henrike Galenkamp,³ Evgeni Levin,^{1,4} Albert K. Groen,¹ Aeilko H. Zwinderman,² and Max Nieuwdorp^{1,5,*}

¹Department of Internal and Vascular Medicine, Amsterdam University Medical Centers, location AMC, 1105 AZ Amsterdam, the Netherlands

²Department of Clinical Epidemiology and Biostatistics, Amsterdam University Medical Centers, location AMC, 1105 AZ Amsterdam, the Netherlands

³Department of Public and Occupational Health, Amsterdam University Medical Centers, location AMC, 1105 AZ Amsterdam, the Netherlands

⁴HorAlzon BV, 2645 LT Delfgauw, the Netherlands

⁵Lead contact

*Correspondence: m.nieuwdorp@amsterdamumc.nl

<https://doi.org/10.1016/j.chom.2022.08.013>

SUMMARY

Previous studies in mainly European populations have reported that the gut microbiome composition is associated with the human genome. However, the genotype-microbiome interaction in different ethnicities is largely unknown. We performed a large fecal microbiome genome-wide association study of a single multi-ethnic cohort, the Healthy Life in an Urban Setting (HELIUS) cohort ($N = 4,117$). Mendelian randomization was performed using the multiethnic Pan-UK Biobank ($N = 460,000$) to dissect potential causality. We identified ethnicity-specific associations between host genomes and gut microbiota. Certain microbes were associated with genotype in multiple ethnicities. Several of the microbe-associated loci were found to be related to immune functions, interact with glutamate and the mucus layer, or be expressed in the gut or brain. Additionally, we found that gut microbes potentially influence cardiometabolic health factors such as BMI, cholesterol, and blood pressure. This provides insight into the relationship of ethnicity and gut microbiota and into the possible causal effects of gut microbes on cardiometabolic traits.

INTRODUCTION

Variation in the human gut microbiome composition has been strongly associated with cardiometabolic health outcomes in numerous studies (Qin et al., 2012; Durack and Lynch, 2019; Reitmeier et al., 2020). The composition of the gut microbiota is governed by a complex array of factors including diet and lifestyle (Muegge et al., 2011; David et al., 2014) but also host genetics. Several genome-wide association studies (GWASs) have identified single-nucleotide polymorphisms (SNPs) that associate with gut microbial abundances (Goodrich et al., 2016; Rothschild et al., 2018). The mechanisms through which the host genome may influence the gut microbiome may involve the immune system functions (Bonder et al., 2016; Wang et al., 2017), pathways involved in glucose metabolism (Wang et al., 2017) and through the metabolic impact of A and B antigens in participants who express these antigens on the surface of mucosal cells (Qin et al., 2022; Sanna et al., 2022). To unravel detailed mechanisms, Mendelian randomization (MR) studies have been used to link potential causal effects of the gut microbiome to clinical phenotypes. For example, recent studies using

MR have underscored that an increase in bacterial butyrate production may improve insulin resistance and that a reduction in production or absorption of propionate can increase type 2 diabetes mellitus (T2DM) risk (Sanna et al., 2019). Another MR study suggested that *Bifidobacterium* may reduce body mass index (BMI) and waist-to-hip ratio (WHR) (Hughes et al., 2020). *Bifidobacterium* could also have a protective effect in ulcerative colitis (Kurilshikov et al., 2021), and another study suggested that the gut microbiome, which was influenced by genotype, could contribute to metabolic syndrome in the host (Lim et al., 2017). Most of these studies however are conducted in populations of European descent. Recently, we and others (Yatsunenko et al., 2012; Martínez et al., 2015; Gupta et al., 2017; Brooks et al., 2018; He et al., 2018) have demonstrated that the gut microbiome differs between participants from different ethnic backgrounds. After adjusting for dietary intake, lifestyle, or sociodemographic factors, up to 3% of the gut microbiome β -diversity was explained by ethnicity alone, suggesting that other variables that are encompassed in ethnicity, such as ancestry, may contribute to the unexplained variance (Deschassaux et al., 2018). Previous studies have already suggested a significant



effect of ethnic ancestry in the associations between the human genome and the gut microbiome (Ishida et al., 2020; Liu et al., 2021; Lopera-Maya et al., 2022; Qin et al., 2022). Currently, no data exist on the interaction between the gut microbiome and host genome in a multi-ethnic population with shared geography, and there is a considerable European bias in genetic studies (Sirugo et al., 2019). We therefore investigated the associations between the human genome and gut microbiome in 4,117 participants from the multi-ethnic Healthy Life in an Urban Setting (HELIUS) cohort and stratified this by six ethnicities to allow for the full genetic variation of our cohort to be represented. Finally, we assessed the causality of gut microbial taxa on cardiometabolic traits in different ethnic groups, using a two-sample MR analysis with outcomes from the Pan-UK Biobank cohort (N = 460,000).

RESULTS

Gut microbiota composition varies between ethnic groups

We performed an analysis of genomic and gut microbiome variation across several ethnic populations sampled from the same geographic location (Figure 1). Confirming our previous results (Deschasaux et al., 2018), now with a doubling of the sample size to 4,117 participants, we observed significant differences in α - and β -diversity, and taxa abundances between ethnicities. This remained after adding relevant confounders (adjusted $p < 0.05$, Figure S1). The South-Asian Surinamese had the lowest α -diversity and the highest incidence of T2DM (29%), whereas the Dutch had the lowest incidence of T2DM (6%) and the highest α -diversity (Table S1). The Shannon index was significantly associated with T2DM status in the HELIUS ($p = 8.13 \times 10^{-17}$, two-sided Wilcoxon rank-sum test). Due to the large variation in the gut microbiome between ethnicities that cannot be explained by known confounding factors, and the mounting evidence of host genome and gut microbiome associations, we next hypothesized that genetic background can contribute to ethnicity-specific variation in the gut microbiome. Furthermore, we hypothesize that the difference in disease burden of ethnic groups sharing a geographic environment may be partly attributed to variations in the gut microbiome, as has also been suggested by other studies (Karlsson et al., 2013; Brooks et al., 2018; Brooks, 2019; Amato et al., 2021).

Host genes strongly associate with gut microbial taxa in an ethnicity-specific manner

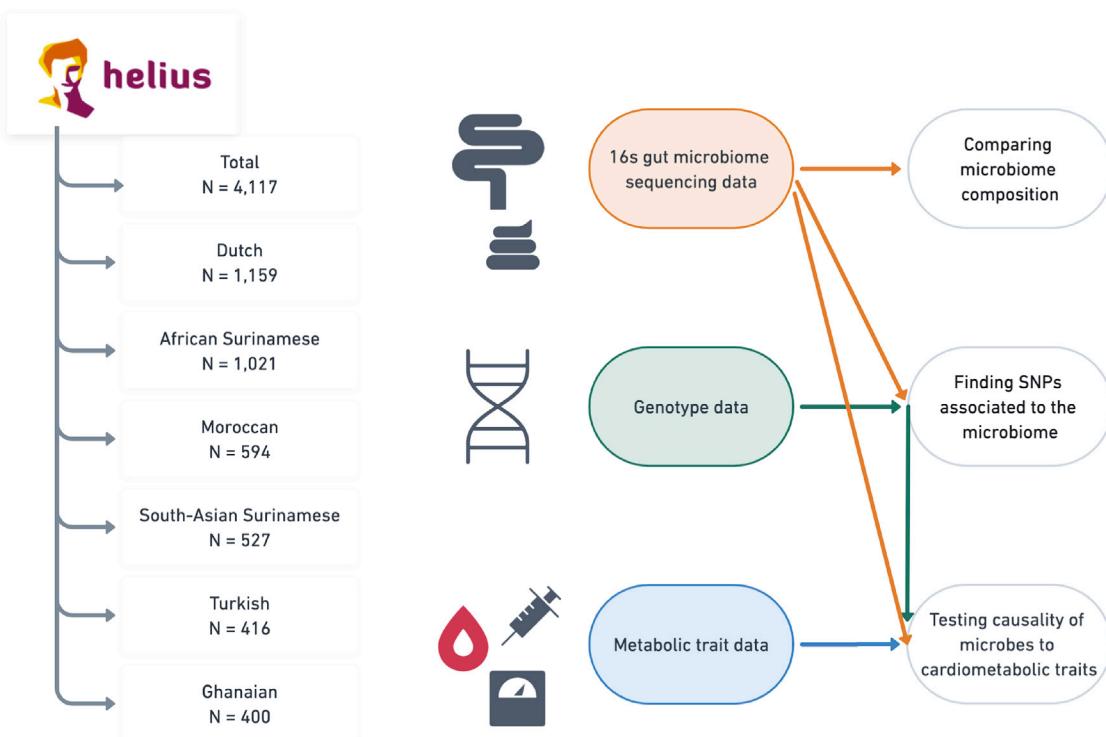
As we saw a clear stratification in genotype between the ethnic groups (Figure S3), we supposed that due to this variation in genetic background and the differences in gut microbiome composition between ethnicities, the heritability of certain microbial taxa may vary per ethnicity. We used genome-wide complex trait analysis (GCTA) to calculate the genetic heritability (h^2) of gut microbial taxa in the “total” cohort, as well as per ethnic group (Table S8). In the total group, 103 taxa were heritable (false discovery rate [FDR] adjusted $p < 0.05$; Figure 2), and across all ethnicities, we found 19 unique microbes that were heritable (adjusted $p < 0.05$). The heritability of the significant microbes in the total group ranged between 0.01 and 0.4, with a mean of 0.06. There were two taxa with a heritability >0.2 in the total

group, namely *Bacteroides uniformis* ($p = 1.2 \times 10^{-6}$, standard error [SE] = 0.05) and *Bacteroidales Prevotellaceae* ($p = 2.16 \times 10^{-5}$, SE = 0.1). Interestingly, these two taxa were some of the most heritable taxa in each ethnic group as well, suggesting that they are highly heritable despite ethnic ancestry. However, the actual heritability estimate of the different ethnic groups varied, with for example *B. uniformis* exhibiting a heritability of 0.64 ($p = 0.046$, SE = 0.18) in the Ghanaian versus a heritability of 0 ($p = 0.5$, SE = 0.12) in the South-Asian Surinamese (Figure S2).

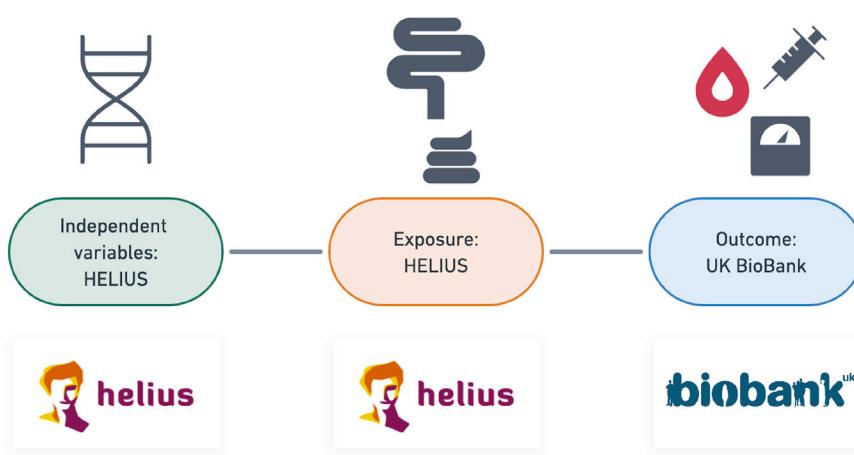
To identify SNPs that are significantly associated with the gut microbiome, we performed a GWAS on the gut microbiome, including α - and β -diversity, enterotypes, microbial abundance or presence, and microbial functional pathways (Figure 1). Due to the technical and geographical homogeneity of our cohort, we performed the GWAS in the total group (quality control and association tests performed in the entire cohort), and per ethnic group with quality control and association tests separately per ethnicity to utilize the full genetic diversity of our cohort and to determine ethnicity-specific signals. We hypothesize that there are ethnicity-specific signals based on two observations: first, due to the large effect of ethnicity on the gut microbiome composition, which cannot be explained by known confounding factors (Deschasaux et al., 2018); second, due to the large variation in human SNPs and their function between ethnicities, for example, the *LCT* gene that is indicative of lactose tolerance in populations of European descent, but not in populations of African descent (Mulcare et al., 2004; Tishkoff et al., 2007).

We found no significant results to the diversity measures or enterotypes at the genome-wide significance threshold ($p > 5 \times 10^{-8}$), which is in concordance with previous studies (Blekhman et al., 2015; Davenport et al., 2015; Bonder et al., 2016; Wang et al., 2016; Goodrich et al., 2016; Turpin et al., 2016; Kolde et al., 2018; Rothschild et al., 2018; Scepanovic et al., 2019; Hughes et al., 2020; Ishida et al., 2020). However, 51 SNPs were significantly associated with gut microbial taxa at the genome-wide significance level ($p < 5 \times 10^{-8}$, Table S3; Figure S4A). Although these were significant at the genome-wide significance level, we additionally calculated a strict study-wide significant threshold, adjusting for all microbial traits analyzed ($p < 1.9 \times 10^{-10}$). None of these associations reached our strict study-wide significant threshold ($p > 1.9 \times 10^{-10}$), which is in line with previous studies reporting only rarely study-wide significant signals (Blekhman et al., 2015; Davenport et al., 2015; Bonder et al., 2016; Wang et al., 2016; Goodrich et al., 2016; Turpin et al., 2016; Kolde et al., 2018; Rothschild et al., 2018; Scepanovic et al., 2019; Hughes et al., 2020; Ishida et al., 2020). We used the genome-wide threshold for exploratory analyses and found that the SNP rs73266725 (in the *FENDRR* gene) and *Dorea formicigenerans* were significantly associated in the total group ($p = 2.89 \times 10^{-8}$, beta = 0.11). *D. formicigenerans* also associated with a SNP (rs8136857, $p = 1.33 \times 10^{-8}$, beta = 0.19) in the *ISX* gene in the South-Asian Surinamese. Interestingly, this microbe belongs to the same genus as *D. longicatena*, which showed the strongest association across all ethnic groups to the SNP rs719400 in the 3-hydroxy-3-methylglutaryl-CoA lyase (*HMGCL*) gene in the Turkish ($p = 8.7 \times 10^{-10}$, beta = -0.3). This was clumped

A



B

**Figure 1. Overview of data and analyses**

(A) Data from 4,117 individuals from the HELIUS cohort with 16S gut microbiome sequencing and genotype data, with no antibiotic use prior to leaving a fecal sample, were included in the study. The aim was to analyze the microbiome composition between ethnicities, identify SNPs associated to the microbiome in each ethnic group, and identify microbial taxa that are potentially causal of cardiometabolic traits. (B) For the MR analysis, the SNPs and microbiome associations are found by a GWAS in the HELIUS cohort. The summary data for clinical outcomes are acquired from the pan GWAS UK Biobank.

with two other SNPs (rs74223776, rs12747718) in the *FUCA1* and *CNR2* loci. In the Moroccan group *Collinsella aerofaciens* ($p = 1.11 \times 10^{-8}$, beta = -0.3) was significantly associated with the intronic SNP rs10409171, in the *LIG1* locus.

The *Bacteroides* genus was associated with multiple SNPs in the African Surinamese. For example, we observed a clump of multiple SNPs (rs73814681, rs11939959, and rs73814654) in the *TEC*, *TXK*, and *NMU* genes, respectively. Many of the

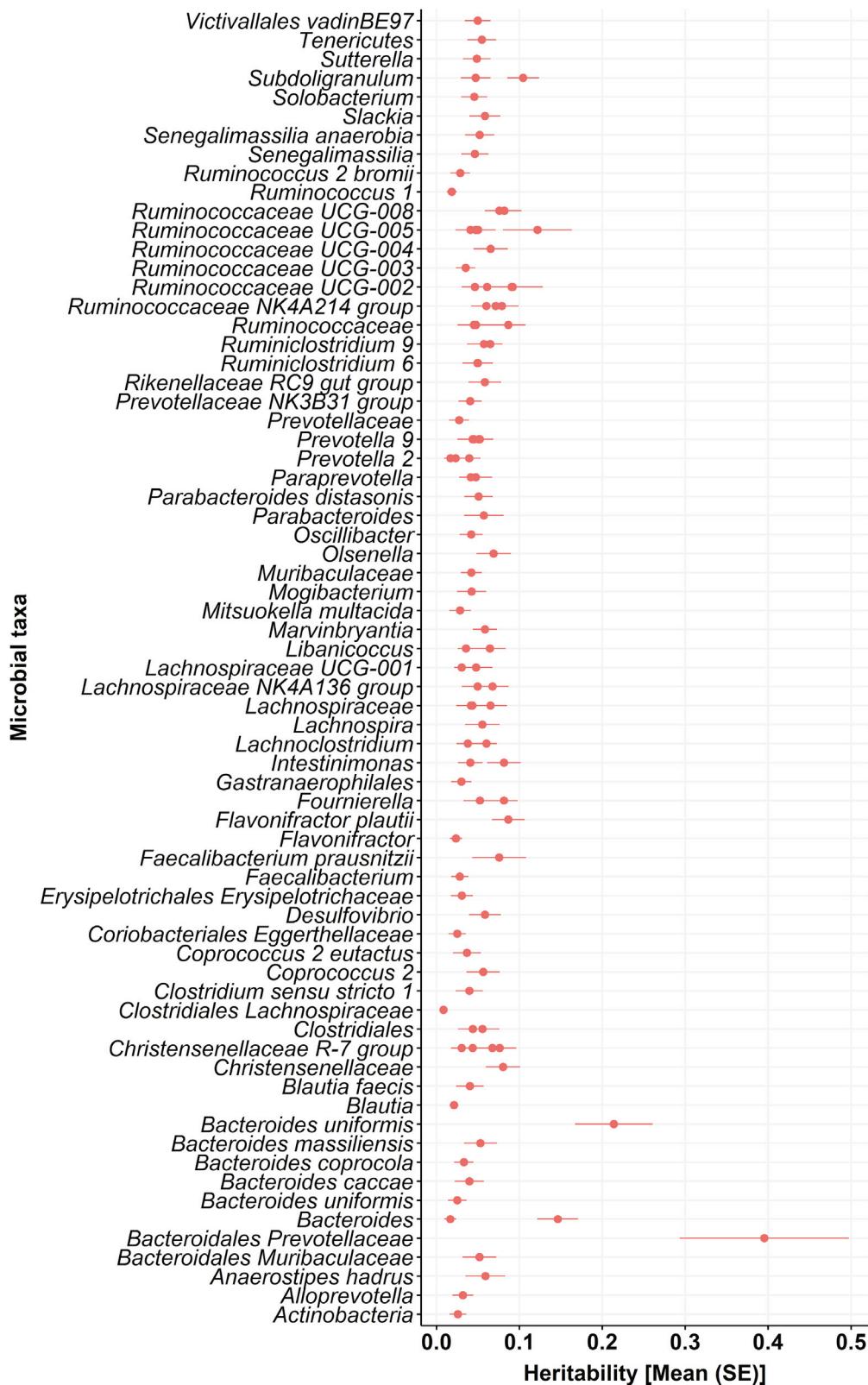
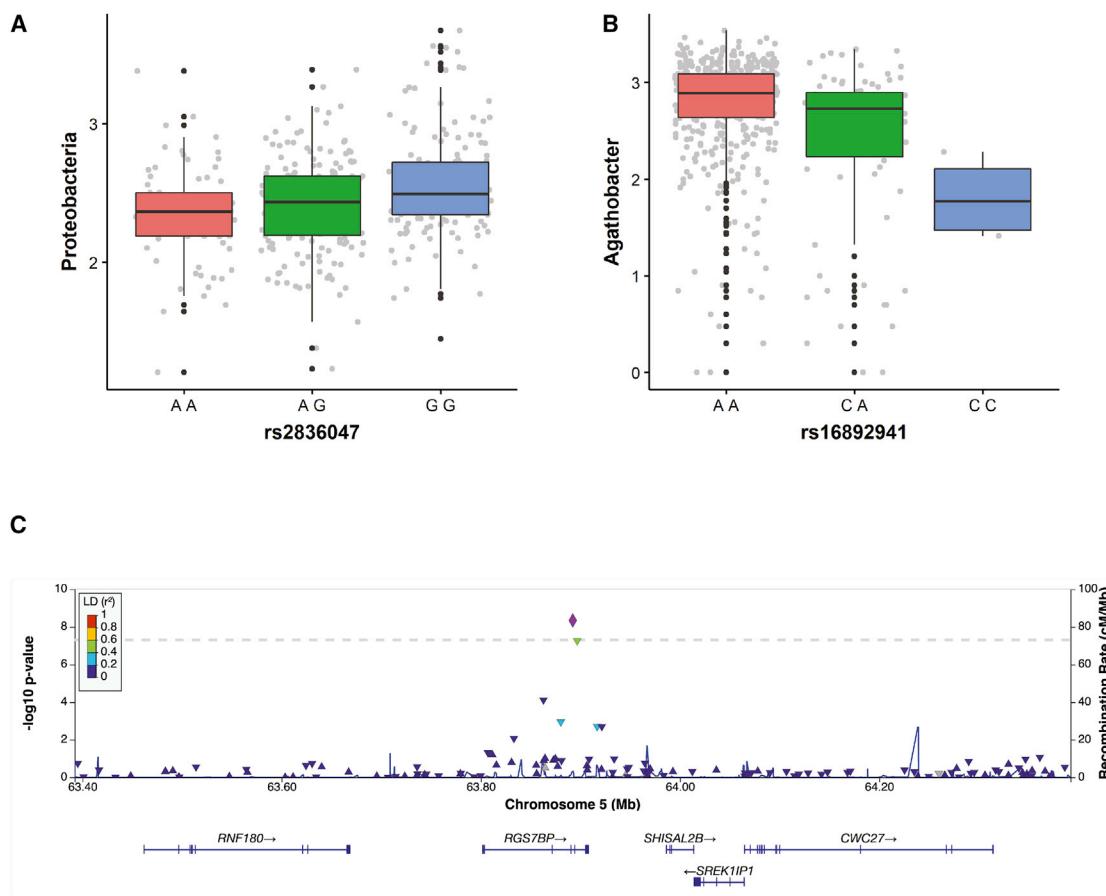


Figure 2. Microbial heritability estimates

Significant (adjusted $p < 0.05$) microbial heritability estimates in the total group. Points represent the mean and lines represent SEs. In the figure, ASVs are collapsed to their closest annotated taxonomy, thus some microbes have multiple heritability estimates.

**Figure 3. Microbes associated with genotype**(A) *Proteobacteria* abundance with rs2836047 in the Ghanaian group.(B) *Agathobacter* abundance with rs16892941 in the South-Asian Surinamese. (A) and (B) are colored by genotype, the center line represents the median, the upper and lower box limits represent quartiles, the whiskers represent 1.5× the interquartile ranges, and outliers are represented by black points.(C) LocusZoom plot of *Agathobacter* with rs16892941 in the Asian reference.

genome-wide significantly associated gut microbes were shared across ethnicities, such as several members of the *Faecalibacterium* genus, which significantly associated with SNPs across all ethnic groups, excluding the Turkish. The SNP rs75039813 in the *GRIK* ($p = 1.85 \times 10^{-9}$, beta = -0.2) locus associated with *Faecalibacterium* in the African Surinamese. *Faecalibacterium* was also associated with the SNP rs7567804 in the *NPAS2* ($p = 4.28 \times 10^{-8}$, beta = -0.15) gene in the Ghanaian. In this group, *Subdoligranulum* was associated with several SNPs, one which was un-annotated (rs7173553) but clumped with rs12050494 in the *SCAMP5* gene ($p = 8.72 \times 10^{-3}$, beta = -0.14). It was also associated with a SNP (rs6799944), which was clumped with two SNPs (rs6787080 and rs147129811) in the *SEC22A* gene ($p = 8.88 \times 10^{-6}$, beta = -0.34 and $p = 6.49 \times 10^{-6}$, beta = -0.34). Moreover, *Agathobacter* was associated with two SNPs, rs16892941 in the *RGS7BP* gene in the South-Asian Surinamese ($p = 1.3 \times 10^{-9}$, beta = -0.41) (Figures 3B and 3C), and rs12677734 in the *CSMD1* gene in the African Surinamese ($p = 3.8 \times 10^{-8}$, beta = -0.26). The SNP rs1241125 in the *OR7K1P* gene, associated with *Clostridiales Ruminococcaceae* ($p = 8.4 \times 10^{-9}$, beta = -0.18) in the Ghanaian. Finally, *Proteobacteria*

was associated with the SNP rs2836047 in *KCNJ6*, a potassium channel ($p = 3.52 \times 10^{-8}$, beta = -0.15), in this group (Figure 3A). Moreover, 11 significant associations were to members of the *Lachnospiraceae* genus.

Replication of SNPs identified in previously reported microbiome GWAS

Moreover, we conducted a replication of previously reported SNP-microbe associations (Table S4) (Davenport et al., 2015; Bonder et al., 2016; Goodrich et al., 2016; Turpin et al., 2016; Wang et al., 2016; Rothschild et al., 2018; Hughes et al., 2020; Kurilshikov et al., 2021; Rühlemann et al., 2021). Of the 330 previously reported SNPs, 75 replicated in this study (adjusted $p < 0.05$, Table S5), among others the *ABO* and *MCM6* loci. The *MCM6* gene and *Bifidobacterium* is one of the most consistent signals in previous gut microbiome GWAS, as it has been replicated in multiple, though not all, published cohorts of mainly European descent (Blekhman et al., 2015; Bonder et al., 2016; Goodrich et al., 2016; Kolde et al., 2018; Rothschild et al., 2018; Hughes et al., 2020). Interestingly, this signal appears to be absent in cohorts of other ethnic backgrounds (Ishida et al.,

2020; Liu et al., 2021). The LCT SNP rs4988235 passed quality control in the Dutch and Ghanaian groups in HELIUS. We successfully replicated the association of *Bifidobacterium* to this SNP in the Dutch ($p = 8.42 \times 10^{-4}$, power = 91%, minor allele frequency [MAF] = 0.26, Figure S5A), which is in line with previous studies and validates our methods (Blekhman et al., 2015; Bonder et al., 2016; Goodrich et al., 2016; Kolde et al., 2018; Rothschild et al., 2018; Hughes et al., 2020). In the Ghanaian group, the association was not significant ($p = 0.59$, power = 33%, MAF = 0.15, Figure S5B). However, there are other SNPs that may be suggestive of lactose tolerance in non-European populations (Tishkoff et al., 2007), but these were either not present, or did not pass quality control in our HELIUS cohort. To expand our search, we identified 30 variants in linkage disequilibrium (LD) with the previously reported SNPs, 11 of which existed in our dataset before quality control. However, they were all filtered out due to low MAF (<5%), except for one variant (rs76337990), which passed the MAF threshold in the Ghanaian group in the HELIUS, but when we did a separate association of this SNP with *Bifidobacterium* in the Ghanaian, no significance was observed ($p = 0.12$).

Dietary intake influences some of the gut microbiome and genome associations

In a subset of participants with dietary data available ($N = 1,053$), we compared the effects of adjusting the microbiome GWAS for dietary intake (Figure S4C). Some associations were influenced by diet, though others remained unchanged. As has been demonstrated previously, many of the reported associations are not necessarily attenuated by diet, though some, such as the LCT and *Bifidobacterium* association, are surely modulated by diet (Lopera-Maya et al., 2022).

Ethnicity-specific host genes are strongly associated with gut microbial function

We also performed a GWAS with PICRUSt2 imputed microbial pathways. Four pathways reached study-wide significance (adjusted $p < 6.8 \times 10^{-10}$): the anhydromuropeptides recycling I pathway (PWY0 1261) to the SNP rs8070080 upstream of the gene *SLFN12* ($p = 2.99 \times 10^{-11}$) in the Ghanaian, the phenylacetate degradation I (PWY0 321) and the 3-hydroxyphenylacetate degradation pathway (3-hydroxyphenylacetate degradation PWY) to the SNP rs11619183 ($p = 4.17 \times 10^{-10}$ and $p = 4.09 \times 10^{-11}$) in the Turkish, and the reductive TCA cycle I (P23 PWY) to the SNP rs17355323 ($p = 2.49 \times 10^{-10}$) in the Moroccan. We found 128 associations below the genome-wide significance threshold ($p < 5 \times 10^{-8}$, Figure S4B; Table S3). The mono-trans, poly-cis decaprenyl phosphate biosynthesis (PWY 6383) pathway associated with the SNP rs78903647 in the *PNPLA6* gene ($p = 3.81 \times 10^{-8}$, beta = 0.28) in the Turkish, whereas the norspermidine biosynthesis pathway (PWY 6562) associated with the SNP rs58348724 in the *GBP5* gene ($p = 2.1 \times 10^{-8}$, beta = 0.3) in the African Surinamese.

Two pathways, the TCA cycle VI (Helicobacter) (RECTICYC) and L-lysine fermentation to acetate and butanoate (P163 PWY), both associated with the SNP rs7100145 in the *PIK3AP1* gene ($p = 1.14 \times 10^{-8}$, beta = -0.22, and $p = 4.98 \times 10^{-8}$, beta = -0.19) in the African Surinamese. Similarly, the pyruvate fermentation to propanoate I (P108 PWY), superpathway of pyridoxal

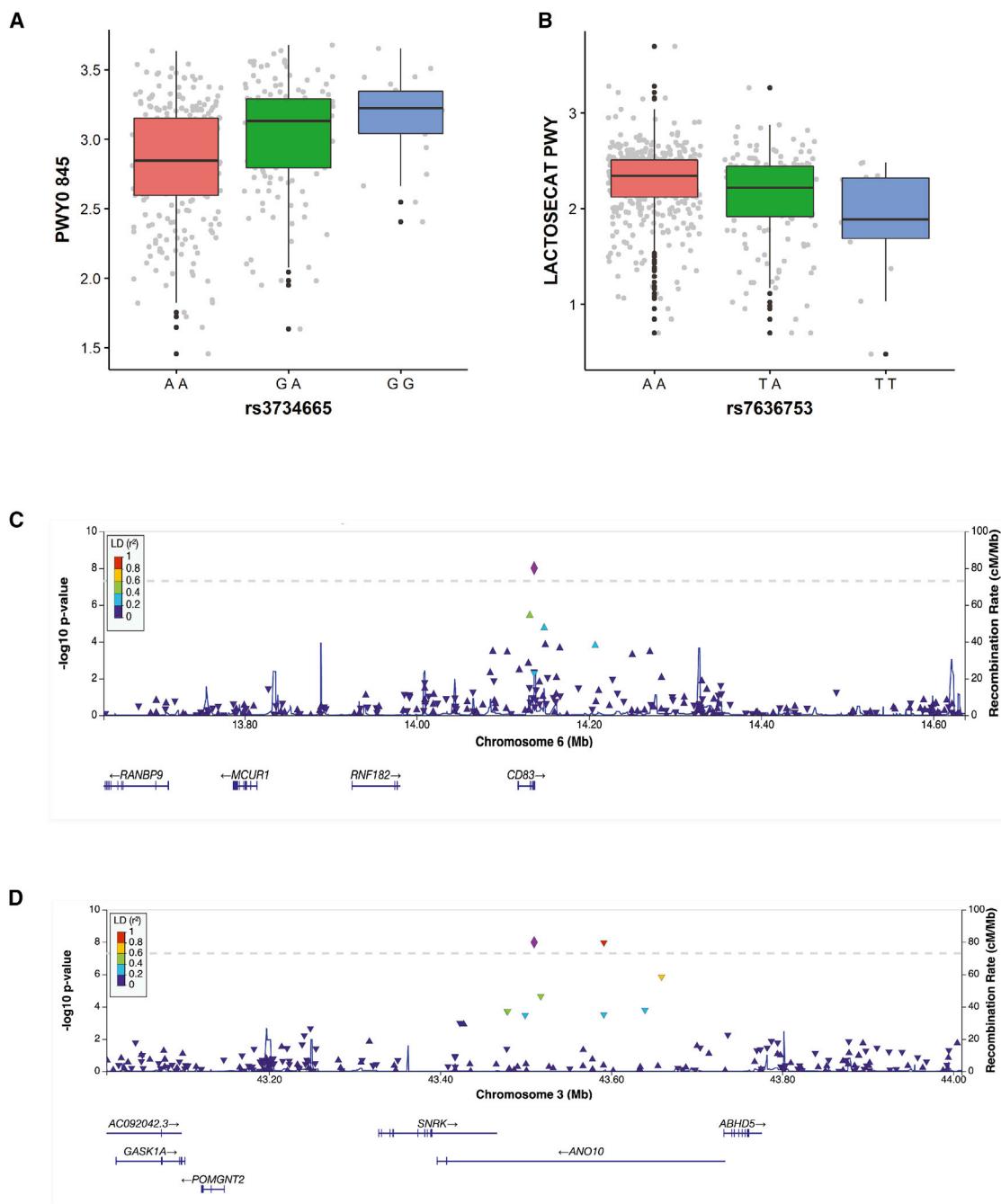
5'-phosphate biosynthesis and salvage (PWY0 845), and pyridoxal 5'-phosphate biosynthesis I (PYRIDOXSYN PWY) pathways all associated with the SNP rs3734665 in the *CD83* gene ($p = 2.99 \times 10^{-8}$, beta = 0.09, $p = 9.95 \times 10^{-9}$, beta = 0.21, and $p = 1.06 \times 10^{-8}$, beta = 0.21) in the Turkish (Figures 4A and 4C). The ethylmalonyl-CoA pathway (PWY 5741) associated with the SNP rs114385359 in the *TFEB* gene ($p = 2.63 \times 10^{-8}$, beta = 0.55) in the Turkish. The inosine 5'-phosphate biosynthesis I (PWY 6123) pathway associated with the SNP rs6970627 in the gene *PON2* ($p = 4.99 \times 10^{-8}$, beta = -0.007) in the total group, whereas the inosine 5'-phosphate degradation pathway (PWY 5695) associated with the SNP rs11863870 in the *PLCG2* gene ($p = 2.8 \times 10^{-8}$, beta = -0.07) in the Ghanaian. Finally, the super pathway of fucose and rhamnose degradation (FUC RHAMCAT PWY) was associated with the SNP rs7186313 in the *PRKCB* gene ($p = 4.5 \times 10^{-9}$, beta = -0.04) in the African Surinamese.

Moreover, in the Dutch, the tetrapyrrole biosynthesis I (from glutamate) (PWY 5188) and the tetrapyrrole biosynthesis II (from glycine) (PWY 5189) pathways both associated with a SNP (rs7612467, $p = 1.49 \times 10^{-8}$, beta = 0.02, and $p = 2.01 \times 10^{-8}$, beta = 0.02) in the *MUC13* gene. The 1,4-dihydroxy-6-naphthoate biosynthesis I pathway (PWY 7374) associated with SNPs (rs111589263 and rs12485354) in the *DYNC1L1* gene ($p = 2.79 \times 10^{-9}$, beta = -0.38, and $p = 6.8 \times 10^{-9}$, beta = -0.36) in the African Surinamese. Two other pathways, super pathway of menaquinol-8 biosynthesis II (PWY 6263) and 1,4-dihydroxy-6-naphthoate biosynthesis II (PWY 7371) also associated with the same SNPs (rs111589263 and rs12485354) in the *DYNC1L1* gene ($p = 7.57 \times 10^{-10}$, beta = -0.32, and $p = 2.12 \times 10^{-8}$, beta = -0.28). The pathway acetyl-CoA fermentation to butanoate (PWY 5676) associated with SNPs in the *MUC5* gene (rs2037089, $p = 1.03 \times 10^{-5}$, beta = -0.02, and rs2943510, $p = 2.07 \times 10^{-6}$, beta = -0.03) in the total group.

Furthermore, the pathway *Bifidobacterium* shunt (P124 PWY) associated with SNPs (rs74224013 and rs74224011) in the *SLC1A7* gene ($p = 4.9 \times 10^{-8}$, beta = -0.2, and $p = 4.96 \times 10^{-8}$, beta = -0.2) in the South-Asian Surinamese. The purine nucleotides degradation II (aerobic) (PWY 6353) pathway associated with SNPs (rs920517 and rs436974) in the *LRRTM4* gene ($p = 4.7 \times 10^{-8}$, beta = -0.03) in the African Surinamese. The lactose and galactose degradation I (LACTOSECAT PWY) pathway associated with several SNPs (rs7636753, rs75962789, rs1842805, and rs6800306) in the *ANO10* gene ($p = 1.04 \times 10^{-8}$, beta = -0.21) in the Moroccan (Figures 4B and 4D). Finally, in the South-Asian Surinamese, the pathway L-glutamate and L-glutamine biosynthesis (PWY 5505) associated with SNPs (rs11808243 and rs11590657) in the *PER3* gene.

Gut microbial taxa and functional pathways as potential causal factors in cardiometabolic traits

Ultimately, we conducted a two-sample MR analysis to investigate potential causal relationships between gut microbes and cardiometabolic traits (Figure 1B). After our strict multiple testing adjustment (FDR $p < 0.05$) only one microbial taxa was significant: the *Clostridiiales Family XIII* was causal in increasing plasma triglycerides in the Central South Asian population (FDR adjusted $p = 3.77 \times 10^{-5}$, beta = 6.42×10^{-2} , Figures 5 and 7A; Table S7). Horizontal pleiotropy was not indicated

**Figure 4. Microbial pathways associated with genotype**

(A) Superpathway of pyridoxal 5'-phosphate biosynthesis and salvage (PWY0 845) abundance with rs3734665 in the Turkish.

(B) Lactose and galactose degradation I (LACTOSECAT PWY) abundance with rs7636753 in the Moroccan. (A) and (B) are colored by genotype, the center line represents the median, the upper and lower box limits represent quartiles, the whiskers represent $1.5 \times$ the interquartile ranges, and outliers are shown as black points.

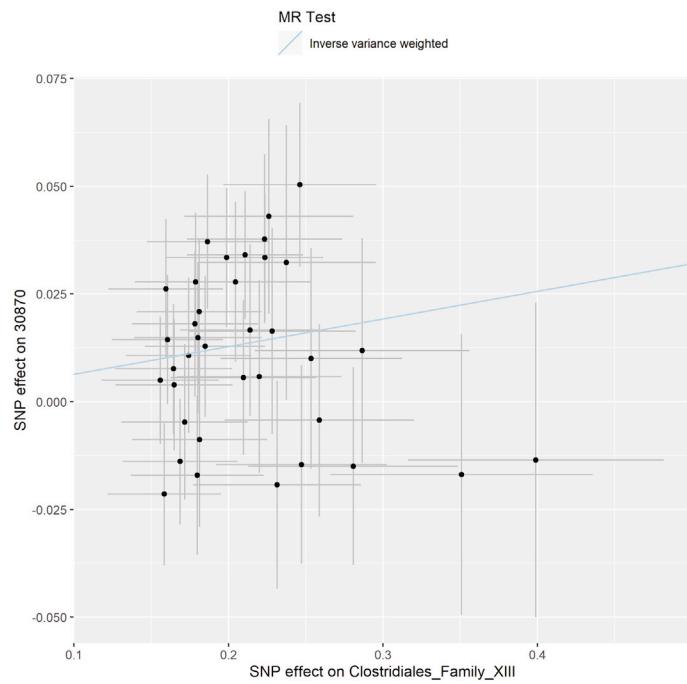
(C) LocusZoom plot of lactose and galactose degradation I (LACTOSECAT PWY) with rs7636753.

(D) LocusZoom plot of pyridoxal 5'-phosphate biosynthesis and salvage (PWY0 845) with rs3734665.

($p > 0.6$), and it was significant with the weighted median ($p = 0.004$), though not with the MR Egger ($p = 0.8$). None of the microbial functional pathways were significant (FDR $p > 0.05$) as causal factors.

Next, we conducted an MR analysis of replicated GWAS results (Table S7). In the total group, 11 associations were significant (adjusted $p < 0.05$). An amplicon sequence variant (ASV) annotated as *Bacteroides eggerthii*, associated with an increased

A



B

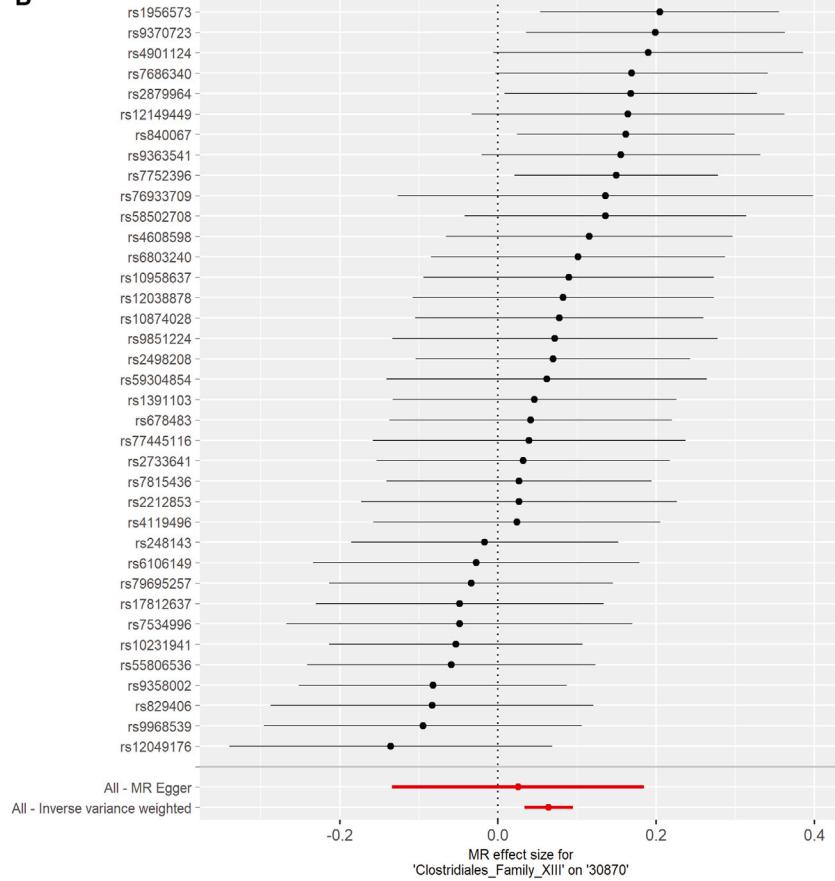


Figure 5. MR results of plasma triglyceride concentration and *Clostridiales Family XIII* abundance in the Central South Asian

(A) Scatter plot: each dot represents the SNP-exposure effect size in standard deviation units with corresponding SE. The blue line represents the average effect as calculated by the inverse variance weighted (IVW) method.

(B) Forest plot: each row represents the SNP-exposure effect size with corresponding SE. The red lines represent the average effect of all SNPs as calculated by MR Egger and the IVW method.

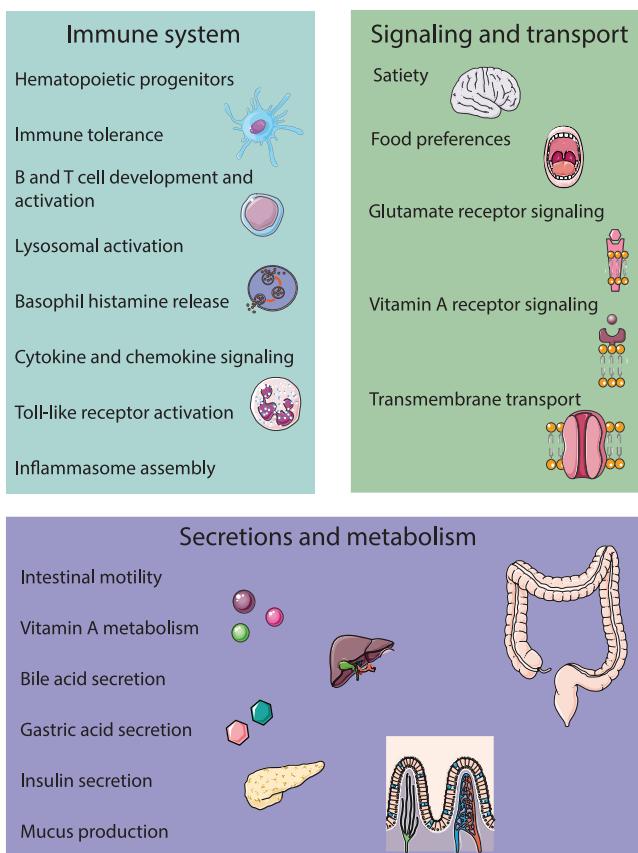


Figure 6. Summary of mechanisms identified in this study of how the human genome associated with the gut microbiome, divided into three larger sets of functions

WHR (FDR $p = 1.13 \times 10^{-2}$, beta = 4.38×10^{-3}), BMI (FDR $p = 1.36 \times 10^{-2}$, beta = 6.63×10^{-3}) and body fat percentage (FDR $p = 4.86 \times 10^{-3}$, beta = 4.69×10^{-3}) (Figure 7B). *Prevotella 2* decreased WHR (FDR $p = 2.23 \times 10^{-4}$, beta = -5.76×10^{-3}), triglycerides (FDR $p = 1.80 \times 10^{-2}$, beta = -6.21×10^{-3}), low-density lipoprotein (LDL) (FDR $p = 2.07 \times 10^{-5}$, beta = -9.63×10^{-3}) and high-density lipoprotein (HDL) (FDR $p = 1.80 \times 10^{-2}$, beta = -6.30×10^{-3}). However, it increased systolic (FDR $p = 2.07 \times 10^{-5}$, beta = 9.42×10^{-3}) and diastolic blood pressure (FDR $p = 3.65 \times 10^{-2}$, beta = 5.56×10^{-3}) (Figure 7C). In the ethnic groups, six results were significant (adjusted $p < 0.05$). Unfortunately, none of the results could be tested for robustness or sensitivity due to the low number of SNPs, likely due to the stringent p value thresholds employed in the validation.

DISCUSSION

Using the largest multi-ethnic cohort with available gut microbiota and genotype data published to date, we identified several host genetic markers that were associated at the genome-wide threshold with gut microbiota. These results may in part contribute to the unexplained variance in microbial composition between the studied ethnic groups. Yet, despite this finding, there appears to be a large number of shared

microbes across ethnic subgroups, which are associated with the human genome. Similar ethnicity-specific genotype associations have been demonstrated for other non-microbe related traits, such as asthma and white blood cell counts (Dahlin et al., 2019; Hu et al., 2021), correspondingly, in this study, we investigated ethnicity-specific genotype to gut microbiome associations. In our study, some of the identified microbe-associated SNPs were in genes known to be involved in the immune system. Moreover, and in line with previous research that suggests involvement of the gut brain axis (Wang et al., 2016; Hughes et al., 2020; Ishida et al., 2020), we also found associations with SNPs in loci that are expressed in the gastrointestinal tract or in the brain. Additionally, we report associations with SNPs in loci that interact with glutamate and mucosa. These results could potentially aid in understanding the difference in cardiometabolic health outcomes between ethnicities. See Figure 6 for a summary of genome and gut microbiome associations.

In the microbiome GWAS in the total group, we found *D. formicigenerans* to be associated with a SNP in the *FENDRR* gene, which reduces the expression of a gene which is important for the differentiation of hematopoietic progenitors (Ono et al., 2005). This suggests that the immune system is able to influence the abundance of certain microbes in the gut. Additionally, we found several ethnicity-specific associations, for example, the association of *D. formicigenerans* with a SNP in the *iSX* gene in the South-Asian Surinamese. This gene is highly expressed in the gastrointestinal tract and encodes a transcription factor that may participate in vitamin A metabolism by regulating *BCO1* gene expression in the intestine (Widjaja-Adhi et al., 2017). Indeed, vitamin A has previously been associated with several microbes and plays an important role in mucin production (Pham et al., 2021). A microbe of the same genus, *D. longicatena*, exhibited the strongest association across all groups to a SNP in the *HMGCL* gene. Mutations in this gene are related to HMGCL deficiency, a condition where the metabolism of dietary fats and the amino acid leucine for energy is disrupted (Essa et al., 2016). *HMGCL* is in the ketogenesis pathway together with D-beta-hydroxybutyrate, which is linked to weight loss, potentially through increasing satiety (Benlloch et al., 2019). This association clumped with the *FUCA1* and *CNR2* loci. *FUCA1* is a part of the mucus regulating genes, which is of high interest considering recent reports from microbiome GWAS on the associations with SNPs in *FUT2* (Kurilshikov et al., 2021; Qin et al., 2022), and the importance of the composition of the mucus layer for the gut microbiome. *CNR2* is expressed in the gastrointestinal tract and modulates inflammatory responses (Izzo, 2004; Wright et al., 2008).

Interestingly, *C. aerofaciens* associated with the *LIG1* locus, which exhibits immune functions (Ellenberger and Tomkinson, 2008; Maffucci et al., 2018). This SNP was reportedly associated with the gene expression of *PLA2G4C* (Zhernakova et al., 2017) and *CARD8*, which cause histamine release of basophils and mediate inflammasome activation (Morita et al., 1983; Linder et al., 2020). *CARD8* has previously been associated with immunity-related diseases such as rheumatoid arthritis (Jenko et al., 2016). Interestingly, *Collinsella* is also believed to be involved in a pro-inflammatory status (Kalinkovich and Livshits, 2019) and associated with diseases such as rheumatoid arthritis

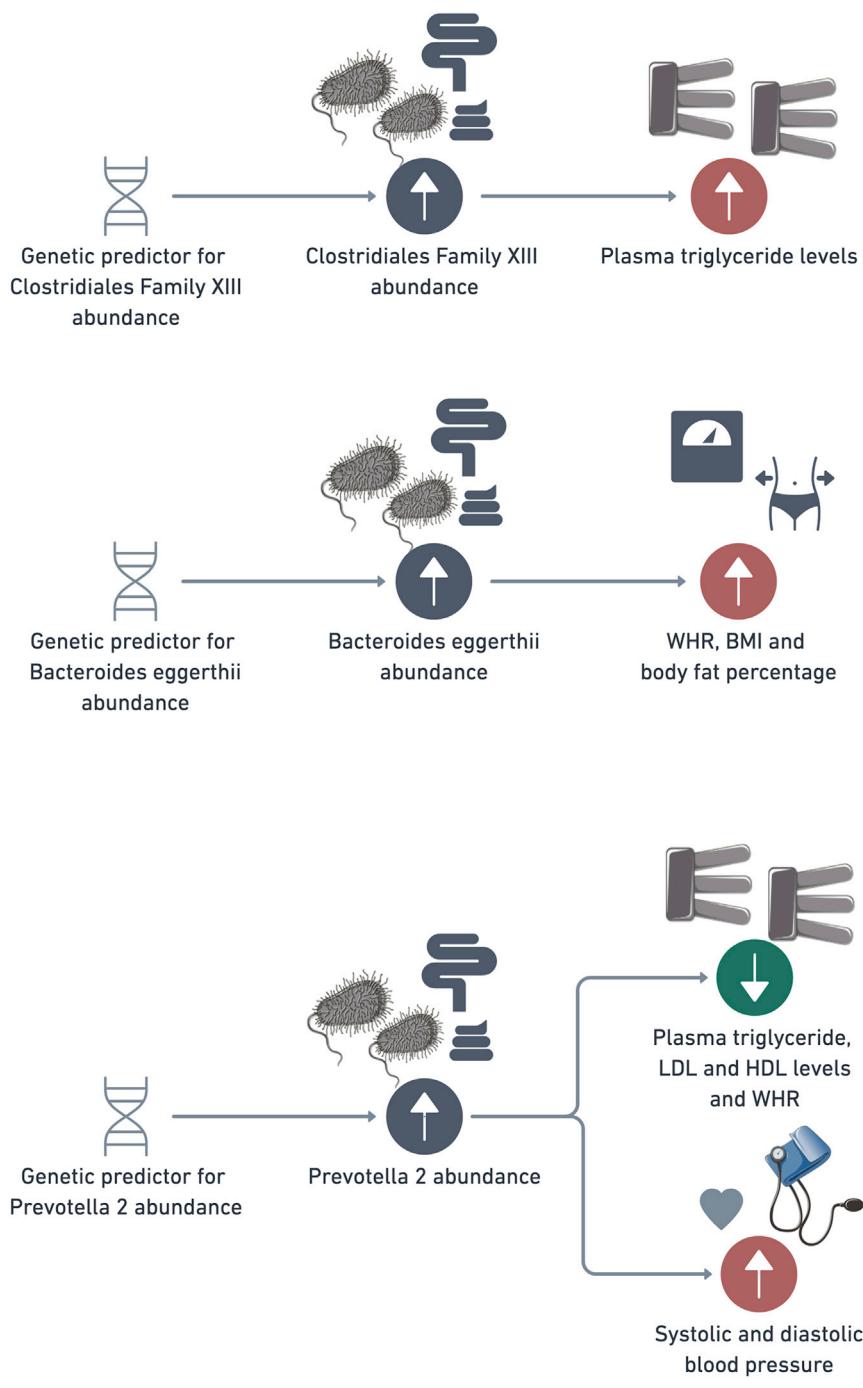


Figure 7. Summary of potential causal mechanisms of how the gut microbiome may influence cardiometabolic health outcomes, identified by MR in this study

tract, influences gastric acid secretion (Brighton et al., 2004), and is associated with a preference for sodium glutamate containing foods (Grippi et al., 2019). Glutamate acts as an important factor in nitrogen metabolism in some *Bacteroides* species (Liu et al., 2017). *Faecalibacterium* associated with the *GRIK2* locus, which encodes a glutamate receptor with high tissue specific expression in the brain (Han et al., 2020). Interestingly, this microbe also associated with a SNP in the *NPAS2* gene, which interacts with vitamin A receptors (Takeda et al., 2011). *Subdoligranulum* was associated with the *SCAMP5* locus, which is involved with exocytosis of cytokines such as *CCL5* (Han et al., 2009). It was also associated with the *SEC22A* locus, which produces an enzyme that is important for vesicle trafficking (Hay et al., 1997). This is an important function in tissues that produce secretions, such as digestive enzymes in the gastrointestinal tract. *Agathobacter* was associated with SNPs in the *CSMD1* and the *RGS7BP* genes, which are expressed with high specificity in the brain (Patil et al., 2018). *CSMD1* is a complement inhibitor and may play an important role in the immune interaction with the microbiome (Kraus et al., 2006).

A SNP in the olfactory receptor *OR7K1P* was negatively associated with *Clostridiales Ruminococcaceae*, which are known butyrate producers (Vital et al., 2017). Interestingly, this gene has previously been associated with increased BMI (Pulit et al., 2019). *Proteobacteria* was associated with the *KCNJ6* locus, encoding a potassium channel which may be involved in the regulation of insulin secretion and is expressed in

(Maeda and Takeda, 2019). *Bacteroides* associated with multiple SNPs; one in the *TEC* locus, which is involved in intracellular signaling mechanisms of cytokine receptors and T cell development and function (Mano, 1999; Boucheron and Ellmeier, 2012); another in the *TXK* locus, which employs a positive regulation of interferon gamma mediated signaling (Takeba et al., 2002); finally, it also associated with the *NMU* locus, which encodes a neuropeptide. This protein plays a role in appetite and feeding regulation, stimulates muscle contractions of the gastrointestinal

the brain and pancreas (Sakura et al., 1995). *Proteobacteria* have previously been implicated in T2DM (Zhang et al., 2021).

Despite the ethnicity-specific results, many of the genome-associated microbes were shared across ethnicities, highlighting the universal effect of the human genome on the gut microbiome. Though many associations do not overlap between ethnic groups, similar functions and loci do associate with related microbes across ethnicities. One example is the SNP rs32122, which associated at the genome-wide level

(5×10^{-8}) with an ASV annotated to *Subdoligranulum* in the total group, and was suggestively associated (5×10^{-5}) with a different ASV, also annotated to *Subdoligranulum*, in the Ghanaian. The lack of shared associations may be due to power and sample size, or indeed a true ethnicity-specific relationship.

Furthermore, we validated some previously reported associations. The *MCM6* gene is in LD with the lactase encoding gene *LCT*, a non-persistent genotype in this gene is associated with an increased abundance of *Bifidobacterium*, possibly due to the capability of bacteria belonging to this genus to utilize lactose (Bonder et al., 2016; Kurilshikov et al., 2021). This association has previously been demonstrated in mainly European descent cohorts, notably not identified in a Japanese or a Chinese cohort (Bonder et al., 2016; Ishida et al., 2020; Kurilshikov et al., 2021; Liu et al., 2021). We validated the *LCT* and *Bifidobacterium* association in our Dutch group, though it was not significant in our Ghanaian group. Similarly, a previous meta-analysis did not find this association in an Israeli cohort, and they found significant heterogeneity for this association between cohorts (Kurilshikov et al., 2021). It was also suggested by a recent study that this signal is likely to vary depending on the ancestry of the cohort studied, due to the variation in genetic determinants for lactose intolerance (Storhaug et al., 2017; Qin et al., 2022). The reasons for this could be a small sample size, low power, or an effect of genetic ancestry, since this gene is reportedly not indicative of lactose tolerance in African descent populations (Mulcare et al., 2004; Tishkoff et al., 2007), and warrants further ethnicity-specific research.

Moreover, the microbiome GWAS on pathways revealed that the anhydromuropeptides recycling I pathway reached study-wide significance in the Ghanaian. The final product of this pathway is a precursor for peptidoglycan biosynthesis, a component of the bacterial cell wall. This pathway was associated with the *SLFN12* locus, which is important for enterocytic differentiation and can be an important target for obesity (Basson et al., 2018; Vomhof-DeKrey et al., 2019). Interestingly, this gene exhibits a high expression in immune cells (Puck et al., 2015) and manifest important immune functions in mice (Geserick, 2004; Li et al., 2012). Additionally, the mono-*trans*, poly-*cis* decaprenyl phosphate biosynthesis, which plays a central role in bacteria cell walls, was significantly associated with the *PNPLA6* locus. This gene produces the neuropathy target esterase protein, which is involved in breakdown of cell membrane lipids (van Tienhoven et al., 2002).

Similar to the results on the microbial taxa GWAS, several microbially expressed pathways were associated with loci that are intimately involved with both the innate and adaptive immune system, such as the norspermidine biosynthesis pathway that associated with the *GBP5* locus, which can activate NLRP3 inflamasome assembly (Shenoy et al., 2012), proinflammatory cytokines, and chemokines (Li et al., 2022). Similarly, the TCA cycle VI (Helicobacter) and L-lysine fermentation to acetate and butanoate pathways both associated with the *PIK3AP1* locus, which contributes to B cell development and is involved in Toll-like receptor (TLR) signaling (Okada et al., 2000; Deason et al., 2018). The pyruvate fermentation to propanoate I, superpathway of pyridoxal 5'-phosphate biosynthesis and salvage, and pyridoxal 5'-phosphate biosynthesis I pathways all associated with the *CD83* locus, which plays an important role in immune toler-

ance (Grosche et al., 2020). The ethylmalonyl-CoA pathway associated with the *TFEB* locus, a transcription factor involved in lysosomal activation (Brady et al., 2018). The inosine 5'-phosphate biosynthesis I pathway associated with the *PON2* locus, which encodes a bacterial quorum sensing mediator and plays a role in the defense response to pathogenic bacteria (Camps et al., 2011; Xiao et al., 2022). On the other hand, the inosine 5'-phosphate degradation pathway was associated with the *PLCG2* locus, which encodes a signaling enzyme that produces diacylglycerol, with important functions in innate immune cells (Jing et al., 2021). The super pathway of fucose and rhamnose degradation associated with the *PRKCB* locus, which is involved in B cell activation (Tsui et al., 2018).

Moreover, several gut microbially expressed pathways were associated with SNPs in genes involved with mucin production, such as the tetrapyrrole biosynthesis I (from glutamate) and the tetrapyrrole biosynthesis II (from glycine) pathways that both associated with the *MUC13* locus, which is highly expressed in the gastrointestinal tract and regulates chemokine secretion (Sheng et al., 2013). The 1,4-dihydroxy-6-naphthoate biosynthesis I pathway, superpathway of menaquinol-8 biosynthesis II and 1,4-dihydroxy-6-naphthoate biosynthesis II all associated with the *DYNC1L1* locus, which is associated with mucins (Chang et al., 2020). The pathway acetyl-CoA fermentation to butanoate associated with the *MUC5* locus, which is also expressed in the gastrointestinal tract (Audie et al., 1993).

Interestingly, similarly to the taxa GWAS, we also found several microbially expressed pathways to be associated with loci involved with glutamate signaling. The pathway *Bifidobacterium* shunt associated with the *SLC1A7* locus, which encodes a glutamate receptor that is also expressed in the intestine (Arriza et al., 1997; Lee et al., 2013), similarly the purine nucleotides degradation II (aerobic) pathway was associated with the *LRRTM4* locus, which acts upstream or within glutamate receptor clustering, and is mainly expressed in the brain (Sinha et al., 2020). We also found the lactose and galactose degradation I pathway associated with several SNPs in the *ANO10* gene, which encodes a chlorine channel that is expressed in the intestine and is involved in transmembrane transport (Chrysanthou et al., 2022). Finally, the pathway L-glutamate and L-glutamine biosynthesis associated with the *PER3* locus, which is a circadian clock gene that may interact with *NPAS2* (Olkkinen et al., 2017), and a polymorphism in *PER3* has been associated with T2DM (Karthikeyan et al., 2014). We identified *NPAS2* to be associated with *Faecalibacterium* in the microbial taxa GWAS.

Moreover, and in line with the microbial taxa level results, several of the pathway-associated SNPs are expressed in brain, but also in the gastrointestinal tract. Additionally, there appears to be a mechanism in which the human genome associates with microbial pathways that are involved in the cellular integrity of microbial cells, or with host mucin, suggesting that human genes may influence the capability of microbes to survive in the gut environment. Our findings corroborate with previous analyses in gut microbiome and genotype interaction studies (Bonder et al., 2016; Goodrich et al., 2016; Wang et al., 2016), underscoring the link between immune system and gut microbiota in humans. Moreover, we reported ethnicity-specific microbiome genotype associations and heritability estimates, which may be linked to differences in disease burden of interest

to an increasingly global society. By analyzing our cohort in an ethnicity-specific manner, we cover a large amount of human genomic variation; only about one third of SNPs were shared across ethnic groups after quality control.

In this study, we additionally aimed to address the potential causality of the gut microbiome on cardiometabolic traits by using MR. Here, we found that the *Clostridiales Family XIII* could increase plasma triglycerides in the Central South Asian population. This bacteria has been implicated in ethnicity-specific associations with insulin resistance in a previous study (Price et al., 2022). In an MR analysis based on replicated SNPs, we found, for example, an ASV annotated as *B. eggerthii*, which associated with an increased WHR, BMI, and body fat percentage. Additionally, *Prevotella 2* was associated with a decreased WHR, triglycerides, LDL, and HDL, but with increased systolic and diastolic blood pressure. Our results are consistent with regard to the direction and biological effects of the microbial traits, underscoring their robustness and reliability, and warrant further investigation in large studies with fecal metagenome sequencing data to elucidate the strain-specific functional capacities that might drive these effects.

Limitations of the study

First, although we present the largest single-cohort multiethnic GWAS to date, the sample size is still relatively small, similar to previous studies on the heritability of the microbiome (Davenport et al., 2015; Goodrich et al., 2016; Rothschild et al., 2018; Hughes et al., 2020; Ishida et al., 2020; Kurilshikov et al., 2021). Traditional GWAS and MR studies often utilize cohorts with hundreds of thousands of participants, thus increasing power and reducing spurious associations. The heterogeneity of the gut microbiota composition also leads to a loss of power. In line with this, as with all studies on the human gut microbiome the sampling variance is large, and the effect of this on the heritability estimate should not be underestimated. This is an issue in all previous studies on microbiome heritability and is likely driven by small sample size. We report associations that reach the genome-wide significance level, as is the norm in previously published gut microbiome GWAS. Due to the small sample sizes and low power, many true signals likely do not reach study-wide significance. Second, we used PICRUSt2 to infer microbial pathway abundance from 16S rRNA sequencing data, which is less accurate than shotgun metagenomic sequencing. Third, the results presented in this study and most previous studies (Sanna et al., 2019; Hughes et al., 2020; Kurilshikov et al., 2020; Rühlemann et al., 2020) lack clear mechanisms by which the gut microbiome casually may influence the outcomes, and thus should only be considered as exploratory leads of potential causal effects. In this regard, we employed a strict p value threshold in the MR analyses. Finally, we used multi-ethnic SNP arrays rather than whole-genome sequencing to increase the number of samples analyzed for an increase in power, though the imputation may still be biased toward populations of European descent. This may further exacerbate an underestimation of ethnicity-specific associations.

Considering the homogeneity in the geography and technical processing and analysis of our multiethnic cohort, in addition to the ability to control for many confounding factors, the effects

are likely to be genuine. Ethnicity was blinded, though not randomized, during fecal sample processing and sequencing; thus, to account for unforeseen effects we included sequencing run as a confounder in our analyses. In the GWAS and MR, we conducted a stratified analysis per ethnicity, thus yielding more homogeneous groups in regard to health status, microbiome composition, dietary intake, and genomic background, as demonstrated in our previous reports in this cohort (Sturkenboom et al., 2016; Deschasaux et al., 2018). There are current efforts to conduct larger studies by combining multiple cohorts to increase power, but the heterogeneity between cohorts may counterbalance an increase in power (Kurilshikov et al., 2021). We believe it could benefit the field if multiethnic studies provide ethnicity-specific analyses rather than only meta-analysis results. This could increase homogeneity, power, and the amount of human genetic variation studied and shed more light on the genetic factors driving the unexplained microbiome variability between ethnicities. Nevertheless, more research in cohorts of non-European descent are greatly needed to further validate our exploratory findings, considering that many microbiome GWAS results do not replicate even between cohorts of European descent (Lopera-Maya et al., 2022; Qin et al., 2022). As the majority of previously published papers on microbiome GWAS and MR studies were conducted in European descent cohorts, our observation that ethnicity may be of great importance highlights the need for future cohort studies aimed at diversifying their populations and including participants of non-European descent.

In conclusion, we demonstrate results that suggest that variations in the gut microbiome and genotype interactions are ethnicity specific, which may contribute to explain the microbiome variability between ethnic groups. Further studies of stratified ethnicity-specific heritability are needed to investigate whether the variations we demonstrate in heritability between the ethnic groups is due to a true biological effect or technical matters such as a difference in sample size. Also, we underscore the usefulness of an MR approach in generating hypothesis for causality testing, which may lead to further insights into the causality of the microbiome in ethnicity-specific cardiometabolic or other phenotypes. Future studies in multiethnic cohorts however are needed to further substantiate our findings.

STAR★METHODS

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

- [KEY RESOURCES TABLE](#)
- [RESOURCE AVAILABILITY](#)
 - Lead contact
 - Materials availability
 - Data and code availability
- [EXPERIMENTAL MODEL AND SUBJECT DETAILS](#)
 - Study population
- [METHOD DETAILS](#)
 - Data collection
 - Genotyping data
 - Stool sample collection
 - Profiling of fecal microbiota composition

- Processing of 16S rRNA gene reads and ASV generation
- Characteristics of gut microbiota composition
- Microbial data preparation
- Dietary patterns

● QUANTIFICATION AND STATISTICAL ANALYSIS

- Statistical analysis of gut microbiota composition between ethnicities
- Testing for SNP association with microbial features
- Testing for SNP association with microbiome beta diversity
- Replication of previously reported SNPs
- Heritability estimates
- Mendelian randomization

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.chom.2022.08.013>.

ACKNOWLEDGMENTS

The HELIUS study is conducted by the Academic University Medical Centers, located in AMC, and the Public Health Service of Amsterdam. Both organizations provided core support for HELIUS. The HELIUS study is also funded by the Dutch Heart Foundation, the Netherlands Organisation for Health Research and Development (ZonMw), the European Union (FP-7), and the European Fund for the Integration of non-EU immigrants (EIF). We gratefully acknowledge the AMC Biobank for their support in biobank management and high-quality storage of collected samples; and we thank the participants of the HELIUS study and its management team, research nurses, interviewers, research assistants, and other staff who have taken part in gathering the data of this study. We would also like to acknowledge input by Willem de Vos and Mary Nicolau on shaping the manuscript. Figures 6 and 7 and the graphical abstract contain illustrations that are adapted from Servier Medical Art (<https://smart.servier.com/>) under a Creative Commons Attribution 3.0 Unported License. The graphical abstract was also created with adapted illustrations from the National Human Genome Research Institute ([genome.gov](https://www.genome.gov)). The study reported here was additionally supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 813781 (on which U.B. is appointed). The funders had no role in the study design; the collection, analysis, and interpretation of data; the writing of the report; or the decision to submit the article for publication. M.N. is funded by a ZonMw VICI grant 2020 (09150182010020).

AUTHOR CONTRIBUTIONS

Conceptualization, supervision, project administration, M.N. Methodology, software, formal analysis, visualization, writing—original draft, U.B. Resources, B.-J.v.d.B., S.-J.P.-S., and H.G. Data curation, U.B., B.F., and H.G. Writing—review and editing, U.B., D.M.B., B.F., B.-J.v.d.B., S.J.P.-S., H.G., E.L., A.K.G., A.H.Z., and M.N. Funding acquisition, B.-J.v.d.B., K.Z., and M.N.

DECLARATION OF INTERESTS

M.N. is on the Scientific Advisory Board of Caelus Pharmaceuticals, the Netherlands. This is not directly relevant to the current paper. There are no patients, products in development, or marketed products to declare.

Received: March 18, 2022

Revised: June 16, 2022

Accepted: August 17, 2022

Published: September 12, 2022

REFERENCES

- Amato, K.R., Arrieta, M.C., Azad, M.B., Bailey, M.T., Broussard, J.L., Bruggeling, C.E., Claud, E.C., Costello, E.K., Davenport, E.R., Dutilh, B.E., et al. (2021). The human gut microbiome and health inequities. *Proc. Natl. Acad. Sci. USA* **118**, e2017947118. <https://doi.org/10.1073/pnas.2017947118>.
- Arriza, J.L., Eliasof, S., Kavanaugh, M.P., and Amara, S.G. (1997). Excitatory amino acid transporter 5, a retinal glutamate transporter coupled to a chloride conductance. *Proc. Natl. Acad. Sci. USA* **94**, 4155–4160. <https://doi.org/10.1073/pnas.94.8.4155>.
- Audie, J.P., Janin, A., Porchet, N., Copin, M.C., Gosselin, B., and Aubert, J.P. (1993). Expression of human mucin genes in respiratory, digestive, and reproductive tracts ascertained by *in situ* hybridization. *J. Histochem. Cytochem.* **41**, 1479–1485. <https://doi.org/10.1177/41.10.8245407>.
- Basson, M.D., Wang, Q., Chaturvedi, L.S., More, S., Vomhof-DeKrey, E.E., Al-Marsoumi, S., Sun, K., Kuhn, L.A., Kovalenko, P., and Kiupel, M. (2018). Schlafen 12 interaction with SerpinB12 and deubiquitylases drives human enterocyte differentiation. *Cell. Physiol. Biochem.* **48**, 1274–1290. <https://doi.org/10.1159/000492019>.
- Benlloch, M., López-Rodríguez, M.M., Cuerda-Ballester, M., Drehmer, E., Carrera, S., Ceron, J.J., Tvarijonaviciute, A., Chirivella, J., Fernández-García, D., and de la Rubia Ortí, J.E. (2019). Satiating effect of a ketogenic diet and its impact on muscle improvement and oxidation state in multiple sclerosis patients. *Nutrients* **11**, 1156. <https://doi.org/10.3390/nu11051156>.
- Beukers, M.H., Dekker, L.H., de Boer, E.J., Perenboom, C.W., Meijboom, S., Nicolaou, M., de Vries, J.H., and Brants, H.A. (2015). Development of the HELIUS food frequency questionnaire: ethnic-specific questionnaires to assess the diet of a multiethnic population in the Netherlands. *Eur. J. Clin. Nutr.* **69**, 579–584. <https://doi.org/10.1038/ejcn.2014.180>.
- Blekhman, R., Goodrich, J.K., Huang, K., Sun, Q., Bukowski, R., Bell, J.T., Spector, T.D., Keinan, A., Ley, R.E., Gevers, D., et al. (2015). Host genetic variation impacts microbiome composition across human body sites. *Genome Biol.* **16**, 191. <https://doi.org/10.1186/s13059-015-0759-1>.
- Bonder, M.J., Kurilshikov, A., Tigchelaar, E.F., Mujagic, Z., Imhann, F., Vila, A.V., Deelen, P., Vatanen, T., Schirmer, M., Smekens, S.P., et al. (2016). The effect of host genetics on the gut microbiome. *Nat. Genet.* **48**, 1407–1412. <https://doi.org/10.1038/ng.3663>.
- Boucheron, N., and Ellmeier, W. (2012). The role of tec family kinases in the regulation of T-helper-cell differentiation. *Int. Rev. Immunol.* **31**, 133–154. <https://doi.org/10.3109/08830185.2012.664798>.
- Brady, O.A., Martina, J.A., and Puertollano, R. (2018). Emerging roles for TFEB in the immune response and inflammation. *Autophagy* **14**, 181–189. <https://doi.org/10.1080/15548627.2017.1313943>.
- Brighton, P.J., Szekeres, P.G., and Willars, G.B. (2004). Neuromedin U and its receptors: structure, function, and physiological roles. *Pharmacol. Rev.* **56**, 231–248. <https://doi.org/10.1124/pr.56.2.3>.
- Brooks, A.W. (2019). How could ethnicity-associated microbiomes contribute to personalized therapies? *Future Microbiol* **14**, 451–455. <https://doi.org/10.2217/fmb-2019-0061>.
- Brooks, A.W., Priya, S., Blekhman, R., and Bordenstein, S.R. (2018). Gut microbiota diversity across ethnicities in the United States. *PLoS Biol* **16**, e2006842. <https://doi.org/10.1371/journal.pbio.2006842>.
- Burgess, S., Foley, C.N., Allara, E., Staley, J.R., and Howson, J.M.M. (2020). A robust and efficient method for Mendelian randomization with hundreds of genetic variants. *Nat. Commun.* **11**, 376. <https://doi.org/10.1038/s41467-019-14156-4>.
- Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., and Holmes, S.P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583. <https://doi.org/10.1038/nmeth.3869>.
- Camps, J., Pujol, I., Ballester, F., Joven, J., and Simó, J.M. (2011). Paraoxonases as potential antibiofilm agents: their relationship with quorum-sensing signals in Gram-negative bacteria. *Antimicrob. Agents Chemother.* **55**, 1325–1331. <https://doi.org/10.1128/AAC.01502-10>.

- Chang, C.C., Chao, K.C., Huang, C.J., Hung, C.S., and Wang, Y.C. (2020). Association between aberrant dynein cytoplasmic 1 light intermediate chain 1 expression levels, mucins and chemosensitivity in colorectal cancer. *Mol. Med. Rep.* 22, 185–192. <https://doi.org/10.3892/mmr.2020.11086>.
- Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4, 7. <https://doi.org/10.1186/s13742-015-0047-8>.
- Chrysanthou, A., Ververis, A., and Christodoulou, K. (2022). ANO10 function in health and disease. *Cerebellum*. <https://doi.org/10.1007/s12311-022-01395-3>.
- Dahlin, A., Sordillo, J.E., Zinitti, J., Iribarren, C., Lu, M., Weiss, S.T., Tantisira, K.G., Lu, Q., Kan, M., Himes, B.E., et al. (2019). Large-scale, multiethnic genome-wide association study identifies novel loci contributing to asthma susceptibility in adults. *J. Allergy Clin. Immunol.* 143, 1633–1635. <https://doi.org/10.1016/j.jaci.2018.11.037>.
- Davenport, E.R., Cusanovich, D.A., Michelini, K., Barreiro, L.B., Ober, C., and Gilad, Y. (2015). Genome-wide association studies of the human gut microbiota. *PLOS One* 10, e0140301. <https://doi.org/10.1371/journal.pone.0140301>.
- David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., et al. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505, 559–563. <https://doi.org/10.1038/nature12820>.
- Deason, K., Troutman, T.D., Jain, A., Challa, D.K., Mandraju, R., Brewer, T., Ward, E.S., and Pasare, C. (2018). BCAP links IL-1R to the PI3K-mTOR pathway and regulates pathogenic Th17 cell differentiation. *J. Exp. Med.* 215, 2413–2428. <https://doi.org/10.1084/jem.20171810>.
- Dekker, L.H., Snijder, M.B., Beukers, M.H., de Vries, J.H., Brants, H.A., de Boer, E.J., van Dam, R.M., Stronks, K., Nicolaou, M., et al. (2011). A prospective cohort study of dietary patterns of non-western migrants in the Netherlands in relation to risk factors for cardiovascular diseases: HELIUS-Dietary Patterns. *BMC Public Health* 11, 441. <https://doi.org/10.1186/1471-2458-11-441>.
- Deschaux, M., Bouter, K.E., Prodan, A., Levin, E., Groen, A.K., Herrema, H., Tremaroli, V., Bakker, G.J., Attaye, I., Pinto-Sietsma, S.J., et al. (2018). Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography. *Nat. Med.* 24, 1526–1531. <https://doi.org/10.1038/s41591-018-0160-1>.
- Douglas, G.M., Maffei, V.J., Zaneveld, J.R., Yurgel, S.N., Brown, J.R., Taylor, C.M., Huttenhower, C., and Langille, M.G.I. (2020). PICRUSt2 for prediction of metagenome functions. *Nat. Biotechnol.* 38, 685–688. <https://doi.org/10.1038/s41587-020-0548-6>.
- Durack, J., and Lynch, S.V. (2019). The gut microbiome: relationships with disease and opportunities for therapy. *J. Exp. Med.* 216, 20–40. <https://doi.org/10.1084/jem.20180448>.
- Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
- Ellenberger, T., and Tomkinson, A.E. (2008). Eukaryotic DNA ligases: structural and functional insights. *Annu. Rev. Biochem.* 77, 313–338. <https://doi.org/10.1146/annurev.biochem.77.061306.123941>.
- Essa, A., Omhani, M., Hamad, S.A.O., Muhammad, S., Ghadeer, A.H., Mohammad, I.S., and Sulman, B. (2016). Recurrent mutation in the HMGCL gene in a family segregating HMG-CoA lyase deficiency. *Int. J. Genet. Mol. Biol.* 8, 11–17. <https://doi.org/10.5897/IJGMB2016.0126>.
- Ferwerda, B., Abdellaoui, A., Nieuwoudt, M., and Zwinderman, K. (2021). A genetic map of the modern urban society of Amsterdam. *Front. Genet.* 12, 727269. <https://doi.org/10.3389/fgene.2021.727269>.
- Geserick, P., Kaiser, F., Klemm, U., Kaufmann, S.H., and Zerrahn, J. (2004). Modulation of T cell development and activation by novel members of the Schlafen (slfn) gene family harbouring an RNA helicase-like motif. *Int. Immunol.* 16, 1535–1548. <https://doi.org/10.1093/intimm/dxh155>.
- Goodrich, J.K., Davenport, E.R., Beaumont, M., Jackson, M.A., Knight, R., Ober, C., Spector, T.D., Bell, J.T., Clark, A.G., and Ley, R.E. (2016). Genetic determinants of the gut microbiome in UK twins. *Cell Host Microbe* 19, 731–743. <https://doi.org/10.1016/j.chom.2016.04.017>.
- Grippi, C., Ahrens, W., Buchecker, K., Chadigeorgiou, C., De Henauw, S., Koni, A.C., Foraita, R., Lissner, L., Molnár, D., Moreno, L.A., et al. (2019). Association between variants of neuromedin U gene and taste thresholds and food preferences in European children: results from the IDEFICS study. *Appetite* 142, 104376. <https://doi.org/10.1016/j.appet.2019.104376>.
- Groot, H.E., van de Vente, Y.J., Verweij, N., Lipsic, E., Karper, J.C., and van der Harst, P. (2020). Human genetic determinants of the gut microbiome and their associations with health and disease: a genome-wide association study. *Sci. Rep.* 10, 14771. <https://doi.org/10.1038/s41598-020-70724-5>.
- Grosche, L., Knippertz, I., König, C., Royzman, D., Wild, A.B., Zinser, E., Sticht, H., Müller, Y.A., Steinkasserer, A., and Lechmann, M. (2020). The CD83 molecule – an important immune checkpoint. *Front. Immunol.* 11, 721. <https://doi.org/10.3389/fimmu.2020.00721>.
- Gupta, V.K., Paul, S., and Dutta, C. (2017). Geography, ethnicity or subsistence-specific variations in human microbiome composition and diversity. *Front. Microbiol.* 8, 1162. <https://doi.org/10.3389/fmicb.2017.01162>.
- Han, C., Chen, T., Yang, M., Li, N., Liu, H., and Cao, X. (2009). Human SCAMP5, a novel secretory carrier membrane protein, facilitates calcium-triggered cytokine secretion by interaction with SNARE machinery. *J. Immunol.* 182, 2986–2996. <https://doi.org/10.4049/jimmunol.0802002>.
- Han, Y., Chen, L., Guo, Y., Wang, C., Zhang, C., Kong, L., and Ma, H. (2020). Class I HDAC inhibitor improves synaptic proteins and repairs cytoskeleton through regulating synapse-related genes in vitro and in vivo. *Front. Aging Neurosci.* 12, 619866. <https://doi.org/10.3389/fnagi.2020.619866>.
- Hay, J.C., Chao, D.S., Kuo, C.S., and Scheller, R.H. (1997). Protein interactions regulating vesicle transport between the endoplasmic reticulum and Golgi apparatus in mammalian cells. *Cell* 89, 149–158. [https://doi.org/10.1016/S0092-8674\(00\)80191-9](https://doi.org/10.1016/S0092-8674(00)80191-9).
- He, Y., Wu, W., Zheng, H.M., Li, P., McDonald, D., Sheng, H.F., Chen, M.X., Chen, Z.H., Ji, G.Y., Zheng, Z.D., et al. (2018). Regional variation limits applications of healthy gut microbiome reference ranges and disease models. *Nat. Med.* 24, 1532–1535. <https://doi.org/10.1038/s41591-018-0164-x>.
- Hemani, G., Tilling, K., and Davey Smith, G. (2017). Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* 13, e1007081. <https://doi.org/10.1371/journal.pgen.1007081>.
- Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., et al. (2018). The MR-Base platform supports systematic causal inference across the human genome. *eLife* 7, e34408. <https://doi.org/10.7554/eLife.34408>.
- Hu, Y., Bien, S.A., Nishimura, K.K., Haessler, J., Hodonsky, C.J., Baldassari, A.R., Highland, H.M., Wang, Z., Preuss, M., Sittani, C.M., et al. (2021). Multi-ethnic genome-wide association analyses of white blood cell and platelet traits in the Population Architecture using Genomics and Epidemiology (PAGE) study. *BMC Genomics* 22, 432. <https://doi.org/10.1186/s12864-021-07745-5>.
- Hua, X., Song, L., Guoqin, Y., Goedert, J.J., Abnet, C.C., Landi, M.T., and Shi, J. (2015). MicrobiomeGWAS: a tool for identifying host genetic variants associated with microbiome composition. Preprint at bioRxiv. <https://doi.org/10.1101/031187>.
- Hughes, D.A., Bacigalupo, R., Wang, J., Rühlemann, M.C., Tito, R.Y., Falony, G., Joossens, M., Vieira-Silva, S., Henckaerts, L., Rymenans, L., et al. (2020). Genome-wide associations of human gut microbiome variation and implications for causal inference analyses. *Nat. Microbiol.* 5, 1079–1087. <https://doi.org/10.1038/s41564-020-0743-8>.
- Ishida, S., Kato, K., Tanaka, M., Odamaki, T., Kubo, R., Mitsuyama, E., Xiao, J.Z., Yamaguchi, R., Uematsu, S., Imoto, S., et al. (2020). Genome-wide association studies and heritability analysis reveal the involvement of host genetics in the Japanese gut microbiota. *Commun. Biol.* 3, 686. <https://doi.org/10.1038/s42003-020-01416-z>.
- Izzo, A.A. (2004). Cannabinoids and intestinal motility: welcome to CB receptors. *Br. J. Pharmacol.* 142, 1201–1202. <https://doi.org/10.1038/sj.bjp.0705890>.

- Jenko, B., Praprotnik, S., Tomšić, M., and Dolžan, V. (2016). NLRP3 and CARD8 polymorphisms influence higher disease activity in rheumatoid arthritis. *J. Med. Biochem.* 35, 319–323. <https://doi.org/10.1515/jomb-2016-0008>.
- Jing, H., Reed, A., Ulanovskaya, O.A., Grigoleit, J.S., Herbst, D.M., Henry, C.L., Li, H., Barbas, S., Germain, J., Masuda, K., et al. (2021). Phospholipase Cγ2 regulates endocannabinoid and eicosanoid networks in innate immune cells. *Proc. Natl. Acad. Sci. USA* 118, e2112971118. <https://doi.org/10.1073/pnas.2112971118>.
- Kalinkovich, A., and Livshits, G. (2019). A cross talk between dysbiosis and gut-associated immune system governs the development of inflammatory arthropathies. *Semin. Arthritis Rheum.* 49, 474–484. <https://doi.org/10.1016/j.semarthrit.2019.05.007>.
- Kamat, M.A., Blackshaw, J.A., Young, R., Surendran, P., Burgess, S., Danesh, J., Butterworth, A.S., and Staley, J.R. (2019). PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics Oxf. Engl.* 35, 4851–4853. <https://doi.org/10.1093/bioinformatics/btz469>.
- Karlsson, F.H., Tremaroli, V., Nookaew, I., Bergström, G., Behre, C.J., Fagerberg, B., Nielsen, J., and Bäckhed, F. (2013). Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* 498, 99–103. <https://doi.org/10.1038/nature12198>.
- Karthikeyan, R., Marimuthu, G., Sooriyakumar, M., BaHammam, A.S., Spence, D.W., Pandi-Perumal, S.R., Brown, G.M., and Cardinali, D.P. (2014). Per3 length polymorphism in patients with type 2 diabetes mellitus. *Horm. Mol. Biol. Clin. Investig.* 18, 145–149. <https://doi.org/10.1515/hmhc-2013-0049>.
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30, 3059–3066. <https://doi.org/10.1093/nar/gkf436>.
- Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. <https://doi.org/10.1093/molbev/mst010>.
- Kolde, R., Franzosa, E.A., Rahnavard, G., Hall, A.B., Vlamakis, H., Stevens, C., Daly, M.J., Xavier, R.J., and Huttenhower, C. (2018). Host genetic variation and its microbiome interactions within the Human Microbiome Project. *Genome Med* 10, 6. <https://doi.org/10.1186/s13073-018-0515-8>.
- Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* 79, 5112–5120. <https://doi.org/10.1128/AEM.01043-13>.
- Kraus, D.M., Elliott, G.S., Chute, H., Horan, T., Pfenninger, K.H., Sanford, S.D., Foster, S., Scully, S., Welcher, A.A., and Holers, V.M. (2006). CSMD1 is a novel multiple domain complement-regulatory protein highly expressed in the central nervous system and epithelial tissues. *J. Immunol.* 176, 4419–4430. <https://doi.org/10.4049/jimmunol.176.7.4419>.
- Kurilshikov, A., Medina-Gomez, C., Bacigalupo, R., Radjabzadeh, D., Wang, J., Demirkan, A., Le Roy, C.I., Garay, J.A.R., Finnicum, C.T., Liu, X., et al. (2020). Genetics of human gut microbiome composition. Preprint at bioRxiv. <https://doi.org/10.1101/2020.06.26.173724>.
- Kurilshikov, A., Medina-Gomez, C., Bacigalupo, R., Radjabzadeh, D., Wang, J., Demirkan, A., Le Roy, C.I., Raygoza Garay, J.A., Finnicum, C.T., Liu, X., et al. (2021). Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nat. Genet.* 53, 156–165. <https://doi.org/10.1038/s41588-020-00763-1>.
- Lee, A., Anderson, A.R., Stevens, M., Beasley, S., Barnett, N.L., and Pow, D.V. (2013). Excitatory amino acid transporter 5 is widely expressed in peripheral tissues. *Eur. J. Histochim.* 57, e11. <https://doi.org/10.4081/ejh.2013.e11>.
- Li, J., and Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 95, 221–227. <https://doi.org/10.1038/sj.hdy.6800717>.
- Li, M., Kao, E., Gao, X., Sandig, H., Limmer, K., Pavon-Eternod, M., Jones, T.E., Landry, S., Pan, T., Weitzman, M.D., et al. (2012). Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. *Nature* 491, 125–128. <https://doi.org/10.1038/nature11433>.
- Li, Y., Lin, X., Wang, W., Wang, W., Cheng, S., Huang, Y., Zou, Y., Ke, J., and Zhu, L. (2022). The proinflammatory role of guanylate-binding Protein 5 in inflammatory bowel diseases. *Front. Microbiol.* 13, 926915. <https://doi.org/10.3389/fmicb.2022.926915>.
- Lim, M.Y., You, H.J., Yoon, H.S., Kwon, B., Lee, J.Y., Lee, S., Song, Y.M., Lee, K., Sung, J., and Ko, G. (2017). The effect of heritability and host genetics on the gut microbiota and metabolic syndrome. *Gut* 66, 1031–1038. <https://doi.org/10.1136/gutjnl-2015-311326>.
- Linder, A., Bauernfried, S., Cheng, Y., Albanese, M., Jung, C., Keppler, O.T., and Hornung, V. (2020). CARD8 inflamasome activation triggers pyroptosis in human T cells. *EMBO J* 39, e105071. <https://doi.org/10.1525/embj.2020105071>.
- Liu, R., Hong, J., Xu, X., Feng, Q., Zhang, D., Gu, Y., Shi, J., Zhao, S., Liu, W., Wang, X., et al. (2017). Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention. *Nat. Med.* 23, 859–868. <https://doi.org/10.1038/nm.4358>.
- Liu, X., Tang, S., Zhong, H., Tong, X., Jie, Z., Ding, Q., Wang, D., Guo, R., Xiao, L., Xu, X., et al. (2021). A genome-wide association study for gut metagenome in Chinese adults illuminates complex diseases. *Cell Discov* 7, 9. <https://doi.org/10.1038/s41421-020-00239-w>.
- Loh, P.R., Danecek, P., Palamara, P.F., Fuchsberger, C., A Reshef, Y., K Finucane, H., Schoenher, S., Forer, L., McCarthy, S., Abecasis, G.R., et al. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* 48, 1443–1448. <https://doi.org/10.1038/ng.3679>.
- Lopera-Maya, E.A., Kurilshikov, A., van der Graaf, A., Hu, S., Andreu-Sánchez, S., Chen, L., Vila, A.V., Gacesa, R., Sinha, T., Collij, V., et al. (2022). Effect of host genetics on the gut microbiome in 7,738 participants of the Dutch Microbiome Project. *Nat. Genet.* 54, 143–151. <https://doi.org/10.1038/s41588-021-00992-y>.
- Lozupone, C., and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71, 8228–8235. <https://doi.org/10.1128/AEM.71.12.8228-8235.2005>.
- Maeda, Y., and Takeda, K. (2019). Host-microbiota interactions in rheumatoid arthritis. *Exp. Mol. Med.* 51, 1–6. <https://doi.org/10.1038/s12276-019-0283-6>.
- Maffucci, P., Chavez, J., Jurkiw, T.J., O'Brien, P.J., Abbott, J.K., Reynolds, P.R., Worth, A., Notarangelo, L.D., Felgentreff, K., Cortes, P., et al. (2018). Biallelic mutations in DNA ligase 1 underlie a spectrum of immune deficiencies. *J. Clin. Invest.* 128, 5489–5504. <https://doi.org/10.1172/JCI99629>.
- Mano, H. (1999). Tec family of protein-tyrosine kinases: an overview of their structure and function. *Cytokine Growth Factor Rev* 10, 267–280. [https://doi.org/10.1016/S1359-6101\(99\)00019-2](https://doi.org/10.1016/S1359-6101(99)00019-2).
- Martínez, I., Stegen, J.C., Maldonado-Gómez, M.X., Eren, A.M., Siba, P.M., Greenhill, A.R., and Walter, J. (2015). The gut microbiota of rural Papua New Guineans: composition, diversity patterns, and ecological processes. *Cell Rep* 11, 527–538. <https://doi.org/10.1016/j.celrep.2015.03.049>.
- McMurdie, P.J., and Holmes, S. (2013). phyloseq: an R Package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8, e61217. <https://doi.org/10.1371/journal.pone.0061217>.
- Mobini, R., Tremaroli, V., Ståhlman, M., Karlsson, F., Levin, M., Ljungberg, M., Sohlin, M., Bertéus Forslund, H., Perkins, R., Bäckhed, F., et al. (2017). Metabolic effects of *Lactobacillus reuteri* DSM 17938 in people with type 2 diabetes: A randomized controlled trial. *Diabetes Obes. Metab.* 19, 579–589. <https://doi.org/10.1111/dom.12861>.
- Morgan, Martin. Dirichlet-Multinomial Mixture Model Machine Learning for Microbiome Data. <https://bioconductor.org/packages/release/bioc/html/DirichletMultinomial.html>.
- Morita, Y., Aida, N., and Miyamoto, T. (1983). Role of phospholipase A2 activation in histamine release from human basophils. *Allergy* 38, 413–418. <https://doi.org/10.1111/j.1398-9995.1983.tb05084.x>.
- Muegge, B.D., Kuczynski, J., Knights, D., Clemente, J.C., González, A., Fontana, L., Henrissat, B., Knight, R., and Gordon, J.I. (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and

- Within humans. *Science* 332, 970–974. <https://doi.org/10.1126/science.1198719>.
- Mulcare, C.A., Weale, M.E., Jones, A.L., Connell, B., Zeitlyn, D., Tarekegn, A., Swallow, D.M., Bradman, N., and Thomas, M.G. (2004). The T Allele of a single-nucleotide polymorphism 13.9 kb Upstream of the lactase gene (LCT) (C–13.9kbT) does not predict or cause the lactase-persistence phenotype in Africans. *Am. J. Hum. Genet.* 74, 1102–1110. <https://doi.org/10.1086/421050>.
- Okada, T., Maeda, A., Iwamatsu, A., Gotoh, K., and Kurosaki, T. (2000). BCAP: the tyrosine kinase substrate that connects B cell receptor to phosphoinositide 3-kinase activation. *Immunity* 13, 817–827. [https://doi.org/10.1016/S1074-7613\(00\)00079-0](https://doi.org/10.1016/S1074-7613(00)00079-0).
- Olkonen, J., Kouri, V.P., Kuusela, E., Ainola, M., Nordström, D., Eklund, K.K., and Mandelin, J. (2017). DEC2 blocks the effect of the ARNTL2/NPAS2 dimer on the expression of PER3 and DBP. *J. Circadian Rhythms* 15, 6. <https://doi.org/10.5334/jcr.149>.
- Ono, R., Nosaka, T., and Hayashi, Y. (2005). Roles of a trithorax Group Gene, MLL, in Hematopoiesis. *Int. J. Hematol.* 81, 288–293. <https://doi.org/10.1532/IJH97.04196>.
- Patil, D.N., Rangarajan, E.S., Novick, S.J., Pascal, B.D., Kojetin, D.J., Griffin, P.R., Izard, T., and Martemyanov, K.A. (2018). Structural organization of a major neuronal G protein regulator, the RGS7-Gβ5-R7BP complex. *eLife* 7, e42150. <https://doi.org/10.7554/eLife.42150>.
- Pham, V.T., Dold, S., Rehman, A., Bird, J.K., and Steinert, R.E. (2021). Vitamins, the gut microbiome and gastrointestinal health in humans. *Nutr. Res.* 95, 35–53. <https://doi.org/10.1016/j.nutres.2021.09.001>.
- Price, C.A., Jospin, G., Brownell, K., Eisen, J.A., Laraia, B., and Epel, E.S. (2022). Differences in gut microbiome by insulin sensitivity status in Black and White women of the National Growth and Health Study (NGHS): A pilot study. *PLoS One* 17, e0259889. <https://doi.org/10.1371/journal.pone.0259889>.
- Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* 5, e9490. <https://doi.org/10.1371/journal.pone.0009490>.
- Puck, A., Aigner, R., Modak, M., Cejka, P., Blaas, D., and Stöckl, J. (2015). Expression and regulation of Schlafin (SLFN) family members in primary human monocytes, monocyte-derived dendritic cells and T cells. *Results Immunol* 5, 23–32. <https://doi.org/10.1016/j.rimim.2015.10.001>.
- Pulit, S.L., Stoneman, C., Morris, A.P., Wood, A.R., Glastonbury, C.A., Tyrrell, J., Yengo, L., Ferreira, T., Marouli, E., Ji, Y., et al. (2019). Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* 28, 166–174. <https://doi.org/10.1093/hmg/ddy327>.
- Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490, 55–60. <https://doi.org/10.1038/nature11450>.
- Qin, Y., Havulinna, A.S., Liu, Y., Jousilahti, P., Ritchie, S.C., Tokolyi, A., Sanders, J.G., Valsta, L., Brożyńska, M., Zhu, Q., et al. (2022). Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort. *Nat. Genet.* 54, 134–142. <https://doi.org/10.1038/s41588-021-00991-z>.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41, D590–D596. <https://doi.org/10.1093/nar/gks1219>.
- Reitmeier, S., Kiessling, S., Clavel, T., List, M., Almeida, E.L., Ghosh, T.S., Neuhaus, K., Grallert, H., Linseisen, J., Skurk, T., et al. (2020). Arrhythmic gut microbiome signatures predict risk of Type 2 diabetes. *Cell Host Microbe* 28, 258–272.e6. <https://doi.org/10.1016/j.chom.2020.06.004>.
- Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P.I., Godneva, A., Kalka, I.N., Bar, N., et al. (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature* 555, 210–215. <https://doi.org/10.1038/nature25973>.
- Rühlemann, M.C., Hermes, B.M., Bang, C., Doms, S., Moitinho-Silva, L., Thingholm, L.B., Frost, F., Degenhardt, F., Wittig, M., Kässens, J., et al. (2021). Genome-wide association study in 8,956 German individuals identifies influence of ABO histo-blood groups on gut microbiome. *Nat. Genet.* 53, 147–155. <https://doi.org/10.1038/s41588-020-00747-1>.
- Sakura, H., Bond, C., Warren-Perry, M., Horsley, S., Kearney, L., Tucker, S., Adelman, J., Turner, R., and Ashcroft, F.M. (1995). Characterization and variation of a human inwardly-rectifying K-channel gene (KCNJ6): a putative ATP-sensitive K-channel subunit. *FEBS Lett* 367, 193–197. [https://doi.org/10.1016/0014-5793\(95\)00498-X](https://doi.org/10.1016/0014-5793(95)00498-X).
- Sanna, S., Kurilshikov, A., van der Graaf, A., Fu, J., and Zhernakova, A. (2022). Challenges and future directions for studying effects of host genetics on the gut microbiome. *Nat. Genet.* 54, 100–106. <https://doi.org/10.1038/s41588-021-00983-z>.
- Sanna, S., van Zuydam, N.R., Mahajan, A., Kurilshikov, A., Vich Vila, A., Vösa, U., Mujagic, Z., Masclee, A.A.M., Jonkers, D.M.A.E., Oosting, M., et al. (2019). Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. *Nat. Genet.* 51, 600–605. <https://doi.org/10.1038/s41588-019-0350-x>.
- Scepanovic, P., et al. (2019). A comprehensive assessment of demographic, environmental and host genetic associations with gut microbiome diversity in healthy individuals. *Microbiome* 7, 557124. <https://doi.org/10.1101/557124>.
- Sheng, Y.H., Triyana, S., Wang, R., Das, I., Gerloff, K., Florin, T.H., Sutton, P., and McGuckin, M.A. (2013). MUC1 and MUC13 differentially regulate epithelial inflammation in response to inflammatory and infectious stimuli. *Mucosal Immunol* 6, 557–568. <https://doi.org/10.1038/mi.2012.98>.
- Shenoy, A.R., Wellington, D.A., Kumar, P., Kassa, H., Booth, C.J., Cresswell, P., and MacMicking, J.D. (2012). GBP5 promotes NLRP3 inflammasome assembly and immunity in mammals. *Science* 336, 481–485. <https://doi.org/10.1126/science.1217141>.
- Sinha, R., Siddiqui, T.J., Padmanabhan, N., Wallin, J., Zhang, C., Karimi, B., Rieke, F., Craig, A.M., Wong, R.O., and Hoon, M. (2020). LRRTM4: a novel regulator of presynaptic inhibition and ribbon synapse arrangements of retinal bipolar cells. *Neuron* 105, 1007–1017.e5. <https://doi.org/10.1016/j.neuron.2019.12.028>.
- Sirugo, G., Williams, S.M., and Tishkoff, S.A. (2019). The missing diversity in human genetic studies. *Cell* 177, 26–31. <https://doi.org/10.1016/j.cell.2019.02.048>.
- Snijder, M.B., Galenkamp, H., Prins, M., Derkx, E.M., Peters, R.J.G., Zwinderman, A.H., and Stronks, K. (2017). Cohort profile: the Healthy Life in an Urban Setting (HELIUS) study in Amsterdam, The Netherlands. *BMJ Open* 7, e017873. <https://doi.org/10.1136/bmjopen-2017-017873>.
- Staley, J.R., Blackshaw, J., Kamat, M.A., Ellis, S., Surendran, P., Sun, B.B., Paul, D.S., Freitag, D., Burgess, S., Danesh, J., et al. (2016). PhenotypeScanner: a database of human genotype-phenotype associations. *Bioinformatics Oxf Engl* 32, 3207–3209. <https://doi.org/10.1093/bioinformatics/btw373>.
- Storhaug, C.L., Fosse, S.K., and Fadnes, L.T. (2017). Country, regional, and global estimates for lactose malabsorption in adults: a systematic review and meta-analysis. *Lancet Gastroenterol. Hepatol.* 2, 738–746. [https://doi.org/10.1016/S2468-1253\(17\)30154-1](https://doi.org/10.1016/S2468-1253(17)30154-1).
- Stronks, K., Snijder, M.B., Peters, R.J., Prins, M., Schene, A.H., and Zwinderman, A.H. (2013). Unravelling the impact of ethnicity on health in Europe: the HELIUS study. *BMC Public Health* 13, 402. <https://doi.org/10.1186/1471-2458-13-402>.
- Sturkenboom, S.M., Dekker, L.H., Lamkaddem, M., Schaap, L.A., de Vries, J.H., Stronks, K., and Nicolaou, M. (2016). Acculturation and dietary patterns among residents of Surinamese origin in the Netherlands: the HELIUS dietary pattern study. *Public Health Nutr* 19, 682–692. <https://doi.org/10.1017/S1368980015001391>.
- Takeba, Y., Nagafuchi, H., Takeno, M., Kashiwakura, J., and Suzuki, N. (2002). Txk, a member of nonreceptor tyrosine kinase of tec family, acts as a Th1 cell-specific transcription factor and regulates IFN-γ gene transcription. *J. Immunol.* 168, 2365–2370. <https://doi.org/10.4049/jimmunol.168.5.2365>.
- Takeda, Y., Kang, H.S., Angers, M., and Jetten, A.M. (2011). Retinoic acid-related orphan receptor γ directly regulates neuronal PAS domain protein 2

- transcription in vivo. *Nucleic Acids Res* 39, 4769–4782. <https://doi.org/10.1093/nar/gkq1335>.
- Thompson, F.E., Subar, A.F., Brown, C.C., Smith, A.F., Sharbaugh, C.O., Jobe, J.B., Mittl, B., Gibson, J.T., and Ziegler, R.G. (2002). Cognitive research enhances accuracy of food frequency questionnaire reports: results of an experimental validation study. *J. Am. Diet. Assoc.* 102, 212–225. [https://doi.org/10.1016/S0002-8223\(02\)90050-7](https://doi.org/10.1016/S0002-8223(02)90050-7).
- Tishkoff, S.A., Reed, F.A., Ranciaro, A., Voight, B.F., Babbitt, C.C., Silverman, J.S., Powell, K., Mortensen, H.M., Hirbo, J.B., Osman, M., et al. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* 39, 31–40. <https://doi.org/10.1038/ng1946>.
- Tsui, C., Martinez-Martin, N., Gaya, M., Maldonado, P., Llorian, M., Legrave, N.M., Rossi, M., MacRae, J.I., Cameron, A.J., Parker, P.J., et al. (2018). Protein kinase C- β dictates B cell fate by regulating mitochondrial remodeling, metabolic reprogramming, and heme biosynthesis. *Immunity* 48, 1144–1159.e5. <https://doi.org/10.1016/j.jimmuni.2018.04.031>.
- Turpin, W., Espin-Garcia, O., Xu, W., Silverberg, M.S., Kevans, D., Smith, M.I., Guttman, D.S., Griffiths, A., Panaccione, R., Otley, A., et al. (2016). Association of host genome with intestinal microbial composition in a large healthy cohort. *Nat. Genet.* 48, 1413–1417. <https://doi.org/10.1038/ng.3693>.
- van Tienhoven, M., Atkins, J., Li, Y., and Glynn, P. (2002). Human neuropathy target esterase catalyzes hydrolysis of membrane lipids. *J. Biol. Chem.* 277, 20942–20948. <https://doi.org/10.1074/jbc.M200330200>.
- Vital, M., Karch, A., and Pieper, D.H. (2017). Colonic butyrate-producing communities in humans: an overview using omics data. *mSystems* 2, e00130-e00117. <https://doi.org/10.1128/mSystems.00130-17>.
- Vomhof-DeKrey, E.E., Lee, J., Lansing, J., Brown, C., Darland, D., and Basson, M.D. (2019). Schlafgen 3 knockout mice display gender-specific differences in weight gain, food efficiency, and expression of markers of intestinal epithelial differentiation, metabolism, and immune cell function. *PLoS One* 14, e0219267. <https://doi.org/10.1371/journal.pone.0219267>.
- Wang, J., Thingholm, L.B., Skiecičienė, J., Rausch, P., Kummen, M., Hov, J.R., Degenhardt, F., Heinsen, F.A., Rühlemann, M.C., Szymczak, S., et al. (2016). Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nat. Genet.* 48, 1396–1406. <https://doi.org/10.1038/ng.3695>.
- Widjaja-Adhi, M.A.K., Palczewski, G., Dale, K., Knauss, E.A., Kelly, M.E., Golczak, M., Levine, A.D., and von Lintig, J. (2017). Transcription factor ISX mediates the cross talk between diet and immunity. *Proc. Natl. Acad. Sci. USA* 114, 11530–11535. <https://doi.org/10.1073/pnas.1714963114>.
- Wright, K.L., Duncan, M., and Sharkey, K.A. (2008). Cannabinoid CB 2 receptors in the gastrointestinal tract: a regulatory system in states of inflammation. *Br. J. Pharmacol.* 153, 263–270. <https://doi.org/10.1038/sj.bjp.0707486>.
- Xiao, Y., Zou, H., Li, J., Song, T., Lv, W., Wang, W., Wang, Z., and Tao, S. (2022). Impact of quorum sensing signaling molecules in gram-negative bacteria on host cells: current understanding and future perspectives. *Gut Microbes* 14, 2039048. <https://doi.org/10.1080/19490976.2022.2039048>.
- Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011>.
- Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., et al. (2012). Human gut microbiome viewed across age and geography. *Nature* 486, 222–227. <https://doi.org/10.1038/nature11053>.
- Ye, Y., and Doak, T.G. (2009). A Parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS Comput. Biol.* 5, e1000465. <https://doi.org/10.1371/journal.pcbi.1000465>.
- Zhang, Z., Tian, T., Chen, Z., Liu, L., Luo, T., and Dai, J. (2021). Characteristics of the gut microbiome in patients with prediabetes and type 2 diabetes. *PeerJ* 9, e10952. <https://doi.org/10.7717/peerj.10952>.
- Zheng, J., Baird, D., Borges, M.C., Bowden, J., Hemani, G., Haycock, P., Evans, D.M., and Smith, G.D. (2017). Recent developments in Mendelian randomization studies. *Curr. Epidemiol. Rep.* 4, 330–345. <https://doi.org/10.1007/s40471-017-0128-6>.
- Zhernakova, D.V., Deelen, P., Vermaat, M., van Iterson, M., van Galen, M., Arindarto, W., van 't Hof, P., Mei, H., van Dijk, F., Westra, H.J., et al. (2017). Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* 49, 139–145. <https://doi.org/10.1038/ng.3737>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human subjects from the HELIUS cohort	HELIUS Executive Board	http://www.heliusstudy.nl/en/over-helius
Deposited data		
16S rRNA gene sequences have been deposited in the European Genome-Phenome Archive	N/A	EGAD00001004106; https://ega-archive.org/
Software and algorithms		
EAGLE2, HAPLOTYPE Reference Consortium (release 1.1), which includes the 1000 Genomes	Loh et al., 2016	https://alkesgroup.broadinstitute.org/Eagle/
PLINK v 2.0	Chang et al., 2015	https://www.cog-genomics.org/plink/2.0/
PLINK v. 1.9	Chang et al., 2015	https://www.cog-genomics.org/plink/1.9/
USEARCH v11.0.667_i86linux64	Edgar, 2010	https://www.drive5.com/usearch/
UNOISE3	N/A	https://drive5.com/usearch/manual/cmd_unoise3.html
SILVA v. 132 reference database	Quast et al., 2013; Callahan et al., 2016	https://www.arb-silva.de/
MAFFT v. 7.427	Katoh, 2002; Katoh and Standley, 2013	https://mafft.cbrc.jp/alignment/software/
FastTree v. 2.1.11	Price et al., 2010	http://www.microbesonline.org/fasttree/
PICRUSt2 v. 2.3.0-b	Douglas et al., 2020	https://huttenhower.sph.harvard.edu/picrust/
minPath	Ye and Doak, 2009	https://omics.informatics.indiana.edu/MinPath/
Dirichlet Multinomial v. 1.28.0	(DirichletMultinomial)	https://bioconductor.org/packages/release/bioc/html/DirichletMultinomial.html
matSpDlite v.1.0	Li and Ji, 2005	https://github.com/snewhouse/BRC_MH_Bioinformatics/blob/master/misc_sh/matSpDlite.R
Phenoscaner v. 2	Staley et al., 2016; Kamat et al., 2019	http://www.phenoscaner.medschl.cam.ac.uk/
microbiomeGWAS v. 1.0	Hua et al., 2015	https://github.com/lncibb/microbiomeGWAS
GCTA-GREML v. 1.93.2	Yang et al., 2011	https://yanglab.westlake.edu.cn/software/gcta/
TwoSampleMR v. 0.5.6	Hemani et al., 2017; Hemani et al., 2018	www.mrbase.org/ https://github.com/MRCIEU/TwoSampleMR
Pan-UK Biobank	N/A	https://pan.ukbb.broadinstitute.org/downloads/index.html
R Statistical Computing Software	The R Foundation	https://www.r-project.org

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact: Max Nieuwdorp, MD, PhD Department of Internal and Vascular Medicine, Amsterdam University Medical Centres, location AMC, 1105 AZ Amsterdam, The Netherlands. e-mail: m.nieuwdorp@amsterdamumc.nl.

Materials availability

This study did not generate new materials.

Data and code availability

- As earlier reported (Deschasaux et al., 2018) the 16S rRNA gene sequences have been deposited in the European Genome-phenome Archive (accession number EGAD00001004106). The HELIUS data are owned by the AMC in Amsterdam, the Netherlands. Any researcher can request the data used in this study by submitting a proposal to the HELIUS Executive Board as outlined at <http://www.heliusstudy.nl/en/researchers/collaboration/>. Collaboration and data that does not include confidential patient information will be provided upon reasonable request. In addition to this, this paper analyzes existing, publicly available data, which are listed in the key resources table.

- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Study population

This study was carried out on a subset of participants from the HELIUS study, which has been previously described ([Stronks et al., 2013](#); [Snijder et al., 2017](#)). The HELIUS study aims to research the causes of unequal burden of disease across different ethnicities. The HELIUS study was conducted with all necessary ethical regulations, and in accordance with the Declaration of Helsinki (6th, 7th revisions); it was approved by the Academic Medical Center (AMC) Medical Ethics Committee, and all participants provided written informed consent. Approximately 25,000 participants aged 18–70 years were recruited from the six largest ethnic groups in Amsterdam, the Netherlands, referred to by their country of origin (not citizenship, which is not used in this study): (1) Dutch: Northwestern European ancestry; (2) Ghanaian: Western African ancestry; (3) Moroccan: Northern African, Mediterranean, and Middle Eastern ancestries; (4) African Surinamese: ‘Creoles’ descending from former slaves, mixed Western African and Dutch ancestries; (5) South-Asian Surinamese: ‘Hindustanis’ descending from workers arriving to the Suriname post-slavery, South-Asian/Indian ancestry; and (6) Turks: Mediterranean, Caucasian, and Middle Eastern ancestries. The final sample size used in this study was 4,117 participants ([Figure 1](#); [Table S1](#), in total 2,175 female and 1,942 male participants), which was based on availability of genetic and microbiome data, after removing participants who had used antibiotics in the three months prior to donating a fecal sample, or where the use of antibiotics was unknown.

METHOD DETAILS

Data collection

After signing the informed consent, participants filled out a questionnaire on sociodemographic characteristics, lifestyle (e.g. dietary habits from a detailed food frequency questionnaire (FFQ)), personal and family history of diseases. Participants were invited to a physical examination where anthropometric characteristics were measured and fasting blood samples were collected. EDTA whole blood and heparin plasma samples were analysed at the AMC Clinical Chemistry Laboratory, and all other samples were transported to the AMC Biobank, where they were stored at –80°C. Regarding ethnicity, a person was defined as of non-Dutch ethnic origin if he/she fulfilled one of two criteria: (1) he/she was born outside the Netherlands and has at least one parent born outside the Netherlands (first generation) or (2) he/she was born in the Netherlands but both parents were born outside the Netherlands (second generation). The Dutch group includes people born in the Netherlands with parents born in the Netherlands. A limitation of the country of birth indicator for ethnicity is that people who are born in the same country might have a different ethnic background, which in the Dutch context is applicable to the Surinamese population. Therefore, participants of Surinamese ethnic origin were grouped by self-reported ethnic origin (obtained by questionnaire) into ‘African,’ ‘South-Asian,’ or ‘other’. The ethnic groups are referred to by their country of origin/ethnic background, without regard to an individual’s actual citizenship.

Genotyping data

The genetic data for the HELIUS study was genotyped at the Erasmus MC Human Genomic Facility, using the Illumina GSA 24v1-0 array, designed for the multiethnic genome-wide content purpose was used. Genotyping of the array was performed with GenomeStudio software according to the in-house protocols of the Human Genomic Facility. Quality control included removing individuals with <95% of the markers genotyped and with a heterozygosity larger than +/- 3 standard deviations. SNPs were filtered based on Minor Allele Frequency (MAF) <1%, and >5% of the individuals not called for the marker and Hardy-Weinberg equilibrium (HWE) violation ($p \leq 0.00001$). Related individuals were excluded using a kinship coefficient of $IBD > 0.185$. The autosomal chromosome data was imputed using the Sanger Imputation Service according to their instructions (<https://imputation.sanger.ac.uk/>). EAGLE2 ([Loh et al., 2016](#)) was used for phasing and the Positional Burrows-Wheeler Transform (PBWT) method using the HAPLOTYPE Reference Consortium (release 1.1), which includes the 1000 Genomes. Imputation SNPs with INFO scores less than 0.3 were removed. When possible, annotation was done with rs numbers, when this was not known the SNPs were annotated with their chromosomal location based on the genome build GRCh37 coordinates. Imputation was done to increase the number of genotypes from the array of approximately 700,000 to approximately 30 million. The ethnicity as described above (Data collection) was thoroughly checked in PCA plots of the HELIUS participants after merging with the 1000 genomes data, as part of the quality control of the genotype data. Thus we have little to no ethnic outliers, and considering the number of participants any effect of potential ethnic outliers on the results of the analyses are deemed to be minor and of low importance. By visually inspecting the principal components (PCs) (calculated on the Total cohort by pca in PLINK v 2.0 ([Chang et al., 2015](#))) we determined that the ten first PCs aligned well with the indicated ethnic groups and were enough to adjust for ethnic background, and that using more than ten would adjust for other factors than ethnicity. Admixture was checked in the African Surinamese population (using ADMIXTURE), and it was deemed that the ten first principal components were able to cover the genetic background of the HELIUS cohort appropriately. The specifics of the admixture analysis of this population are however beyond the scope of this paper. In addition to the initial quality control of the genotyping data, subsequent quality control on the imputed genotyping data was performed using PLINK v. 1.9 ([Chang](#)

et al., 2015), to ensure fewer and more accurate SNPs to be used for the association analysis. SNPs with a MAF <5% were removed. SNPs were pruned for linkage disequilibrium (LD) with parameters –indep-pairwise 50 5 0.5. SNPs were filtered by Hardy-Weinberg equilibrium (HWE) violation $p < 0.00001$. This was done for the Total dataset, and separately for all ethnicity groups as well as the diet subset, mainly to ensure accurate MAF levels, but also to cover a large amount of human genomic variation (Table S2). The final number of SNPs after quality control was approximately 1 million per group, which is suitable with a genome-wide significance threshold of 5×10^{-8} , see Table S2 for the exact number of SNPs and number of individuals per group.

Stool sample collection

Each participant collected one stool sample using a collection tube, and they were asked to bring it to the research location within 6 hours after collection. If this was not possible they were asked to bring the frozen sample to the research location the morning after storing it in their freezer overnight. The samples were temporarily stored at -20°C at the research location, before daily transportation to the AMC where they were stored at -80°C after being checked by a nurse. Information regarding probiotic use, antibiotic use in the past three months, and if they had diarrhea in the past week, was asked when the participant handed in the sample.

Profiling of fecal microbiota composition

Library preparation and sequencing was performed at the Wallenberg Laboratory (Sahlgrenska Academy at the University of Gothenburg, Sweden). Using a repeated bead beating method, total genomic DNA was extracted from a 150 mg fecal sample aliquot as previously described (Mobini et al., 2017). The V4 region of the 16S rRNA gene was sequenced on a MiSeq system (RTA v. 1.17.28, bundled with MCS v. 2.5; Illumina) with 515F and 806R primers designed for dual indexing (Kozich et al., 2013) and the v2 Illumina kit (2 \times 250bp paired-end reads). Amplification of 16S rRNA genes from each sample were conducted in duplicate reactions in volumes of 25 μ l containing 1 \times Five Prime Hot Master Mix (5PRIME GmbH), 200 nM of each primer, 0.4 mg ml $^{-1}$ BSA, 5% dimethylsulfoxide, and 20 ng of genomic DNA. Polymerase chain reaction (PCR) was performed as follows: initial denaturation for 47 minutes at 94 °C; 25 cycles of denaturation for 45 seconds at 94 °C; annealing for 60 seconds at 52 °C; elongation for 90 seconds at 72 °C; final elongation for 10 minutes at 72 °C. Duplicate samples were merged, and thereafter purified and quantified with the NucleoSpin Gel and PCR Clean-Up kit (Macherey-Nagel), and the Quant-iT PicoGreen dsDNA kit (Invitrogen), respectively. The purified PCR products were diluted to 10 ng/ μ l and pooled in equal amounts. To remove short amplification products, the pooled amplicons were purified a second time using Ampure magnetic purification beads (Agencourt). Each sample had a corresponding negative control, where gel electrophoresis was used to confirm the absence of detectable PCR products. Positive controls were not included though mock samples were used to optimize the protocol. The pooled amplicons were mixed with PhiX control DNA (Illumina) to prepare libraries for sequencing. The input DNA had a concentration of 3 pM and contained 15% PhiX, and resulted in the generation of approximately 700K clusters/mm 2 . All analytical procedures were blinded for ethnicity (but not randomized).

Processing of 16S rRNA gene reads and ASV generation

The raw sequencing reads were processed using USEARCH (v11.0.667_i86linux64 (Edgar, 2010)). For paired-end merging, ‘max-diffs’ of 30 was used in the overlapping region for the merging step (using ‘fast_mergepairs’ command) and ‘fastq_maxee’ of 1 was used as a quality filter threshold (using the ‘fastq_filter’ command). After merging paired-end reads and quality filtering, remaining contigs were dereplicated and unique sequences were denoised using the UNOISE3 algorithm, in order to obtain Amplicon Sequence Variants (ASVs). All merged reads were subsequently mapped against the resulting ASVs to produce an ASV table. ASVs not matching expected amplicon length were filtered out (i.e. ASV sequences longer than 260 basepair or shorter than 250 basepair). Taxonomy was assigned with the ‘assignTaxonomy’ function from the ‘dada2’ R package (v 1.12.1) and the SILVA (v. 132) reference database (Quast et al., 2013; Callahan et al., 2016). ASV sequences were then aligned using MAFFT (v. 7.427; Katoh, 2002; Katoh and Standley, 2013) with default settings. A phylogenetic tree was constructed from the resulting multiple sequence alignment with the ‘double precision’ build of FastTree (v. 2.1.11; Price et al., 2010) using a generalized time-reversible model (-gtr). The ASV table, taxonomy and tree were integrated using the ‘phyloseq’ R package (v. 1.28.0; McMurdie and Holmes, 2013). The ASV table was rarefied to 14,932 counts per sample using the ‘vegan’ R package (v 2.5-6). Of 6,056 sequenced samples, 24 had insufficient counts (<5,000 counts per sample) and were excluded at the rarefaction stage. The complete dataset thus contained 6,032 samples and 22,532 ASVs. Participants that had used antibiotics in the past 3 months before the fecal sample were excluded (including participants where this information was missing), as well as participants not registered as one of the 6 main ethnicities (Dutch: N = 1,159, South-Asian Surinamese N = 527, African Surinamese N = 1,021, Ghanaian N = 400, Moroccan N = 594, Turkish N = 416.) The final total data set comprised 4,117 participants. PICRUSt2 (v. 2.3.0-b; Douglas et al., 2020) with default settings was used to generate pathway level information from the 16S sequencing data.

Characteristics of gut microbiota composition

The dissimilarities in gut microbiota composition between individuals (β -diversity) were calculated with the Bray–Curtis dissimilarity index, as well as weighted and unweighted UniFrac distance calculated at the ASV level (function ‘vegdist’ vegan v. 2.5-6 R package for Bray–Curtis and function ‘UniFrac’ phyloseq v 1.30.0 R package for weighted and unweighted UniFrac; Lozupone and Knight, 2005). Bray–Curtis dissimilarities were plotted using unconstrained PCoA (function pcoa ape v. 5.4-1 R package). The α -diversity

of gut microbiota for each individual was assessed with several indices calculated at the ASV level: richness (number of unique ASVs: function ‘estimate’ vegan R package); Shannon, Simpson, and Inverse Simpson index: function diversity ‘vegan’ R package; and PD (Faith’s PD: function pd ‘picante’ v. 1.8.2 R package).

Microbial data preparation

The microbial data was summarized to phylum, family, genus or species level, filtered to only keep taxa with >20 counts in >5% of participants (Figure S10). Since some microbes can be redundant across different taxonomic levels we used the findCorrelation function (R caret package v. 6.0-90) to remove highly correlated taxa (based on Spearman’s correlation coefficient >0.9). This method identifies variables with high correlation and removes the one with the highest mean absolute correlation. This approach was chosen to avoid a very strict multiple-testing correction, though it may lead to potential false negatives, as suggested previously (Sanna et al., 2019; Hughes et al., 2020). This resulted in 401 taxa in the final analysis (237 ASVs, 60 species, 21 families, 76 genera, 7 phyla). The pathway data as output from PICRUSt2 was filtered to keep only pathways with >20 counts in >5% of participants. For any pair of pathways that were correlated with a Spearman’s correlation coefficient >0.9, only one was kept. This resulted in 169 pathways for further analysis. For any pathway or taxa where <5% of participants had zero abundance, a pseudo count of 1 was added and the data was \log_{10} transformed (function ‘filter_taxa’ and ‘transform_sample_counts’ phyloseq v. 1.30.0 R package). If >5% of participants had zero abundance, the data was instead encoded into a presence/absence pattern and analysed as a binary trait to combat the zero-skewed distribution of their abundance. Enterotypes were calculated in the microbiome data of all participants combined. It was done on the rarefied ASV table, which had not been filtered or transformed. First all ASVs were aggregated to genus level. The ASVs that could not be aggregated to a genus (i.e. where the ASV was annotated as NA at the genus level) were combined into one column of non-annotated genera. This was not included for the clustering. The clustering was done using the Dirichlet Multinomial package (Ye and Doak, 2009; DirichletMultinomial) in R, with 2, 3 or 4 clusters. Three clusters were selected based on the laplace goodness of fit, and each subject was thus assigned in one of three clusters. By visually inspecting the top driving genera in each cluster, the clusters were annotated as *Bacteroides*, *Prevotella*, and *Ruminococcus*. Each enterotype cluster was encoded in a one-hot-encoding manner to allow for further analysis in plink.

Dietary patterns

Of the sample size which had genetic and microbiome data ($N = 4,117$), detailed dietary information in the form of FFQs from the HELIUS-Dietary Patterns study (Dekker et al., 2011; Beukers et al., 2015; Sturkenboom et al., 2016) were available for 1,053 samples from five main ethnicities (no participants of Ghanaian descent were present). The dietary intake was recorded through an ethnicity-specific semi-quantitative FFQ with a reference period of 4 weeks. To accurately assess the dietary intake of the participants the FFQs were designed for the different populations studied (Beukers et al., 2015). The Dutch FFQ was designed using cognitive theory (Thompson et al., 2002), which laid the basis for the ethnicity specific FFQs. This means that the layout and main food items were similar across ethnicities, but the FFQ was further adjusted based on food intake data per ethnic group. Several methods were applied to identify potential problems with filling in the questionnaires: expert judgement by 13 dieticians of the specific ethnic origins; 30 interviews conducted with participants with a focus on comprehensibility; 9 focus groups consisting of 6–12 participants of the same ethnic origin; and cognitive interviews among 7 Moroccan participants. Dietary patterns were calculated as four principal component scores based on the FFQ data reported per subject, using principal(nfactors=4,rotate='varimax') from the psych package v. 1.9.12.31 in R, as previously suggested (Deschasaux et al., 2018).

QUANTIFICATION AND STATISTICAL ANALYSIS

All analyses in R were conducted using v. 3.6.1.

Statistical analysis of gut microbiota composition between ethnicities

All analyses were adjusted for the following covariates: age, sex, BMI, proton pump inhibitor (PPI) use, metformin use, 16S rRNA sequencing run and educational level. Alpha diversity indices were compared between ethnicities using ‘lm’ in R, the alpha diversity measure was adjusted for covariates and residuals were extracted for a final ‘lm’ analysis of ethnicity. Beta diversity was tested with a PERMANOVA (‘adonis’, in R (Distance ~ age + sex + BMI + PPI use + metformin use + 16S rRNA sequencing run + educational level + ethnicity, permutations = 1000)). Residuals of transformed taxa abundances were computed using ‘lm’ in R ($\log_{10}(\text{Microbe abundance}+1)$ ~ age + sex + BMI + PPI use + metformin use + 16S rRNA sequencing run + educational level). These residuals were compared between ethnicities with ‘lm’ in R (Residuals of adjusted microbe abundance ~ Ethnicity), and a multiple testing adjusted p-value (p.adjust with Bonferroni method in R) <0.05 was deemed significant. Analyses were conducted on data from the participants where all information on covariates were present ($N=4,115$, two samples had missing information for BMI).

Testing for SNP association with microbial features

The statistical association test was done using PLINK v. 2.0, using the parameters: -glm and –covar-variance-standardize. The data sets were all analyzed with covariates (top 10 genetic principal components (PCs) calculated by pca in PLINK v. 2.0, age, sex, BMI, PPI use, metformin use, 16S rRNA sequencing run and educational level). Ten genetic PCs were chosen as appropriate adjustment for genetic ancestry, both based on visual inspection of the PCA plot as well as advised previously in multiethnic cohorts (Kurilshikov

et al., 2021). In addition to this we verified that the genomic inflation factor did not exceed 1.1 in any of the resulting associations. The genomic inflation factor was calculated as follows in R: The chi-square statistic was calculated as $\text{chisq} = (\text{Z-score})^2$. Genomic inflation factor = $\text{median}(\text{chisq})/\text{qchisq}(0.5,1)$, i.e. the median of the observed chi-squared test statistics was divided by the expected median of the corresponding chi-squared distribution. To determine the study-wide statistical significance threshold in this study we used matSpDlite v.1.0 (Li and Ji, 2005) to evaluate the number of independent tests using eigenvalue variance. This is based on the fact that microbes interact in an independent manner in their environment, as suggested by a recent study (Qin et al., 2022). The number of independent tests was 268 for 401 taxa, resulting in a Bonferroni adjusted study wide significance threshold of $(5 \times 10^{-8})/268$, i.e. 1.9×10^{-10} . The same method was applied to estimate the study-wide significance threshold for the microbial functional pathway GWAS, which resulted in 74 independent tests for 169 pathways. Using this information the study-wide significance threshold for the microbial pathways was $(5 \times 10^{-8})/74$, i.e. 6.8×10^{-10} . A genome-wide significance threshold was set to 5×10^{-8} , to allow for comparison with previous studies. The p-value threshold used for the alpha and beta diversity measures was 5×10^{-8} . The annotations of resulting SNPs were done with Phenoscanner v. 2 (<http://www.phenoscanner.medschl.cam.ac.uk/>, (Staley et al., 2016; Kamat et al., 2019)). We conducted the analyses in each of the six ethnic groups, as well as in the Total cohort. Since the quality control of the genotyping data was conducted in each group (and the Total group) separately, we do not believe genetic differences would cause false positives. As this is one single cohort with all the data collected and analyzed in an identical manner, we argue that the grouping of all ethnicities into one total, as opposed to conducting a meta-analysis, yields the most power (>4,000 participants) with low risk of false positives due to heterogeneity or technical factors. The approach of pooling samples that were genotyped and processed in an identical manner, and conducting ancestry-combined as well as ancestry specific GWAS, has been demonstrated elsewhere (Dahlin et al., 2019; Hu et al., 2021). In a previous study conducted in the HELIUS cohort we compared the effects of a meta-analysis versus pooling the samples into one group, and found that a pooled analyses appears to be appropriate in hypothesis generating analyses in a cohort of this size (Ferwerda et al., 2021). In addition, we used dietary principal components (described above in the section Dietary patterns) as additional confounders in a GWAS. This was compared to a GWAS in the same subset of participants, but without diet covariates. This was not stratified by ethnicity as the number of participants were too small. The genotype data in this subset of participants underwent quality control separately from the other subsets. In all following analyses, we did not adjust for diet, as this is typically not done in the microbiome GWAS field. In addition to this we were also interested in identifying potential diet driven effects, such as the LCT SNP association with *Bifidobacterium* abundance which is believed to be largely driven by dairy intake, or potential taste signaling receptor SNPs or SNPs related to food intolerances (e.g. lactose or gluten intolerance). In order to find these types of potential associations, we thus deemed it inappropriate to correct for diet. Power was calculated with the function power.calc.linear from the genpwr R package version 1.0.4. The True.Model and Test.Model were both set to Additive, and the sample size, MAF, and effect size were obtained from the results from PLINK.

Testing for SNP association with microbiome beta diversity

The package microbiomeGWAS (v. 1.0; Hua et al., 2015) was used to test for associations between beta diversity (measured by Bray-Curtis distance, weighted UniFrac, or Unweighted UniFrac) and genotype, with covariates (top 10 genetic PCs calculated by pca in PLINK v. 2.0, age, sex, BMI, PPI use and metformin use).

Replication of previously reported SNPs

SNPs that were previously reported as significantly associated with a microbial feature ($p < 5 \times 10^{-8}$) were tested for association in our HELIUS cohort (see Table S4 for the originally reported SNPs and the SNP in the closest genomic position in the HELIUS cohort. The significant replications are reported in Table S5). The SNPs from HELIUS that were in the closest genomic location to the previously reported SNPs were chosen to use for replication. Note that as the ethnic groups can be considered independent and since the quality control was done separately, each group contributes with their own set of closest SNPs. The cut-off p-value threshold for significance was below 0.05/330, per group.

Heritability estimates

GCTA-GREML (v. 1.93.2; Yang et al., 2011) was used to calculate chip heritability for all microbial traits. This was done both for the Total cohort as well as for the separate ethnic groups. The first 10 genetic principal components (as calculated by plink v. 2.0) were used as covariates. The genetic relationship matrix was calculated for each group separately, using the quality controlled genotype data as prepared with plink. The microbial traits were either encoded as Presence/Absence or Abundance, see methods section *Microbial data preparation* for details. The results are presented in Table S8, including the encoding of the microbial traits.

Mendelian randomization

SNP selection

Microbes that were significantly associated with SNPs at the genome-wide level were used as exposure for MR ($p < 5 \times 10^{-8}$). SNPs that were significantly associated ($p < 5 \times 10^{-5}$) with any of these microbial features in the GWAS were used as instrumental variables. This approach to increase the number of instrumental variables has been demonstrated by previous studies to be a viable methodology in microbiome MR (Sanna et al., 2019; Kurihikov et al., 2021; Rühlemann et al., 2021). If the SNP association to the exposure (microbial feature) results in an F statistic > 10 ($F \text{ statistic} = \beta^2/\text{SE}^2$) the SNP was approved for downstream analysis. For the binary microbial traits, the explained variance was calculated using the get_or_from_lor function from the TwoSampleMR R package

(v. 0.5.6; [Hemani et al., 2017](#); [Hemani et al., 2018](#)). The F statistic was then calculated as $\frac{r^2*(N-1-k)}{(1-r^2)*k}$, where r^2 is the explained variance, N is the sample size, and k is the number of instrumental variables. Two sample MR analysis was performed using the TwoSampleMR R package (<https://github.com/MRCIEU/TwoSampleMR>) v. 0.5.5, which is created and provided by MR-Base (www.mrbase.org/). The GWAS summary-level outcome data was extracted from the Pan-UK Biobank website: (<https://pan.ukbb.broadinstitute.org/downloads/index.html>) for the following outcomes: triglycerides (UK Biobank code 30870), BMI (UK Biobank code 21001), glucose (UK Biobank code 30740), cholesterol (UK Biobank code 30690), HDL (UK Biobank code 30760), LDL (UK Biobank code 30780), waist to hip ratio (UK Biobank code whr), body fat percentage (UK Biobank code 23099), type 2 diabetes (UK Biobank code E11), diastolic blood pressure (UK Biobank code 4079), systolic blood pressure (UK Biobank code 4080). This data was imported as outcome data using the TwoSampleMR R package, using either the 'Meta' results (corresponding to the results from the MR analysis in HELIUS for the Total group), or the 'AFR' results (corresponding to the exposure from the GWAS in HELIUS for the Ghanaian and African Surinamese participants), or the 'CSA' results (corresponding to the exposure from the GWAS in HELIUS for the South-Asian Surinamese participants), or the 'EUR' results (corresponding to the exposure from the GWAS in HELIUS for the Dutch and Turkish participants), or the 'MID' results (corresponding to the exposure from the GWAS in HELIUS for the Moroccan participants). If only one SNP was available the causality was tested using Wald ratio, as this is a generally accepted and widely used method among multiple different research groups for single SNP MR analyses ([Groot et al., 2020](#); [Hughes et al., 2020](#); [Rühleman et al., 2021](#)). If there were two SNPs the IVW method was used as an extension of the Wald ratio estimates for multiple SNPs, considering it is a standard MR method with high power ([Zheng et al., 2017](#); [Burgess et al., 2020](#)). If there were 3 or more SNPs, the IVW method was used, along with MR Egger and weighted median to assess robustness. Results with a Benjamini-Hochberg adjusted p-value <0.05 (p.adjust with method='fdr' in R, calculated per group) are reported in [Table S7](#).