Original research

Host and gut microbial tryptophan metabolism and type 2 diabetes: an integrative analysis of host genetics, diet, gut microbiome and circulating metabolites in cohort studies

Qibin Qi ,[1,2] Jun Li ,[2,3] Bing Yu,[4] Jee-Young Moon,[1] Jin C Chai,[1] Jordi Merino ,[5,6] Jie Hu,[7] Miguel Ruiz-Canela,[8,9] Casey Rebholz,[10] Zheng Wang ,[1] Mykhaylo Usyk,[11] Guo-Chong Chen,[1] Bianca C Porneala,[12] Wenshuang Wang,[4,13] Ngoc Quynh Nguyen,[4] Elena V Feofanova,[4] Megan L Grove,[4] Thomas J Wang,[14] Robert E Gerszten,[15,16] Josée Dupuis,[17] Jordi Salas-Salvadó,[9,18] Wei Bao,[19] David L Perkins,[20] Martha L Daviglus,[20] Bharat Thyagarajan,[21] Jianwen Cai,[22] Tao Wang,[1] JoAnn E Manson,[3,23] Miguel A Martínez-González,[8,9] Elizabeth Selvin,[10] Kathryn M Rexrode,[7] Clary B Clish,[24] Frank B Hu,[25] James B Meigs,[6,12] Rob Knight,[26] Robert D Burk,[1,11] Eric Boerwinkle,[4] Robert C Kaplan[1,27]

► Additional online supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2021-324053).

For numbered affiliations see end of article.

Correspondence to
Dr Qibin Qi, Department of Epidemiology and Population Health, Yeshiva University Albert Einstein College of Medicine, Bronx, New York, USA; qibin.qi@einsteinmed.org

QQ, JL and BY contributed equally.

Received 7 January 2021
Accepted 7 June 2021

To cite: Qi Q, Li J, Yu B, *et al*. *Gut* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/gutjnl-2021-324053

ABSTRACT

Objective Tryptophan can be catabolised to various metabolites through host kynurenine and microbial indole pathways. We aimed to examine relationships of host and microbial tryptophan metabolites with incident type 2 diabetes (T2D), host genetics, diet and gut microbiota.

Method We analysed associations between circulating levels of 11 tryptophan metabolites and incident T2D in 9180 participants of diverse racial/ethnic backgrounds from five cohorts. We examined host genome-wide variants, dietary intake and gut microbiome associated with these metabolites.

Results Tryptophan, four kynurenine-pathway metabolites (kynurenine, kynurenate, xanthurenate and quinolinate) and indolelactate were positively associated with T2D risk, while indolepropionate was inversely associated with T2D risk. We identified multiple host genetic variants, dietary factors, gut bacteria and their potential interplay associated with these T2D-relaetd metabolites. Intakes of fibre-rich foods, but not protein/tryptophan-rich foods, were the dietary factors most strongly associated with tryptophan metabolites. The fibre-indolepropionate association was partially explained by indolepropionate-associated gut bacteria, mostly fibre-using *Firmicutes*. We identified a novel association between a host functional *LCT* variant (determining lactase persistence) and serum indolepropionate, which might be related to a host gene-diet interaction on gut *Bifidobacterium*, a probiotic bacterium significantly associated with indolepropionate independent of other fibre-related bacteria. Higher milk intake was associated with higher levels of gut *Bifidobacterium* and serum indolepropionate only among genetically lactase non-persistent individuals.

Conclusion Higher milk intake among lactase non-persistent individuals, and higher fibre intake were associated with a favourable profile of circulating tryptophan metabolites for T2D, potentially through the host–microbial cross-talk shifting tryptophan metabolism toward gut microbial indolepropionate production.

Significance of this study

What is already known on this subject?

- Tryptophan can be catabolised to various metabolites through host kynurenine and microbial indole pathways.
- Evidence from animal studies suggests a host–microbiota interaction on tryptophan metabolism which may affect host metabolic health.
- Circulating levels of some tryptophan metabolites have been associated with risk of type 2 diabetes in human studies.
- Genetic variants located on genes that are involved in the host tryptophan-kynurenine pathway and dietary factors have been associated with circulating tryptophan metabolites, but the role of gut microbiome and its interplay with host genetics and diet in tryptophan metabolism remain unclear in humans.

INTRODUCTION

Tryptophan is an essential amino acid that plays a critical role in human health and disease.[1] In addition to its role in serotonin and melatonin biosynthesis, tryptophan is the sole source for the

Significance of this study

What are the new findings?

- In large-scale populations with diverse racial/ethnic backgrounds, circulating levels of tryptophan and several kynurenine-pathway metabolites were positively associated with risk of type 2 diabetes, while a microbial indole derivative, indolepropionate, was inversely associated with risk of type 2 diabetes. The indolepropionate-type 2 diabetes association was suggested to be potentially causal by the latent causal variable model.
- Intakes of fibre-rich foods, but not protein/tryptophan-rich foods, were the dietary factors most strongly associated with circulating tryptophan metabolites. The fibre-indolepropionate association can be partially explained by indolepropionate-associated gut bacteria (mostly fibre-utilising *Firmicutes* bacteria).
- We identified a novel genetic association between a host functional *LCT* variant (determining lactase persistence) and serum indolepropionate, which might be a result of host gene–diet interaction on gut *Bifidobacterium*. Higher milk intake was associated with higher levels of gut *Bifidobacterium* and serum indolepropionate only among genetically lactase non-persistent individuals.

How might it impact on clinical practice in the foreseeable future?

- These findings contribute to our understanding of the host–microbial cross-talk in tryptophan metabolism and its implications in human metabolic health and disease, and may help to identify high-risk individuals based on circulating metabolite profiles for targeted interventions through dietary intervention and gut microbiota modification.

kynurenine pathway (online supplemental figure S1),[1] in which tryptophan is first catabolised into kynurenine, mainly regulated by indoleamine 2,3-dioxygenase (IDO) and trypophan-2,3-dioxygenase (TDO) and then kynurenine is processed into several downstream metabolites, including kynurenate, xanthurenate and quinolinate. The kynurenine pathway is involved in immune activation and inflammation regulation,[1] and has been associated with obesity and insulin resistance.[2 3] In addition, tryptophan can be catabolised by gut microbiota, producing a variety of indole derivatives (eg, indoleacetate, indolelactate and indolepropionate) which have been shown to have beneficial effects on host metabolism.[4]

Emerging evidence from animal studies suggests a host–microbiota interaction on tryptophan metabolism which may affect host metabolic health.[5] In mice with genetic deficiency of IDO, tryptophan metabolism may shift from the host kynurenine pathway towards gut microbial indole derivative production, leading to an improvement in insulin sensitivity.[5] In human studies, metabolomics using a broad-spectrum of metabolites found that plasma levels of tryptophan[6] and two kynurenine-pathway metabolites (kynurenate and xanthurenate)[7] were associated with increased risk of type 2 diabetes (T2D), while a microbial metabolite of tryptophan, indolepropionate, was associated with decreased risk of T2D,[8] but relationships of other tryptophan metabolites with T2D remains unclear. Genome-wide association studies (GWAS) of the human blood metabolome identified genetic loci associated with some tryptophan metabolites and many of them might be involved in host tryptophan-kynurenine metabolism or metabolite transportation.[9–11] Dietary tryptophan is the only source of tryptophan and its catabolites for humans,[1] while several human studies found strong positive associations of fibre-rich food (eg, fruits and vegetables) and fibre intake with circulating indolepropionate levels.[8 12 13] The human gut microbiome might be involved in this relationship but underlying mechanisms remain unclear, since no evidence has shown that indolepropionate can be derived from microbial catabolism of phytochemical compounds or fibre fermentation. A recent study in women reported an association between gut microbiome composition and serum indolepropionate which appeared to be independent of host dietary fibre intake.[14] To the best of our knowledge, no studies have examined host and microbial tryptophan metabolism and T2D integrating data on host genome-wide variants, dietary intake, gut microbiome and circulating levels of both host and microbial tryptophan metabolites. There is a need to integrate different layers of data to identify more relevant associations, and more importantly, potential links among these association signals, which may help better understand host–microbial cross-talk in tryptophan metabolism and its implication in human metabolic health.

In this study, we hereby examined prospective associations between circulating levels of 11 major host and microbial tryptophan metabolites and incident T2D in five epidemiological cohorts of multiple racial/ethnic groups, hypothesising that kynurenine-pathway metabolites are associated with higher risk of T2D, while microbial indole derivatives are associated with lower risk of T2D. Furthermore, by integrating multiomics data, we identified host genetic, dietary and gut microbial factors associated with these metabolites.

METHODS

Study population

The main study population was the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), with subsequent replication analyses conducted in four additional cohorts of multiple racial/ethnic groups: the Atherosclerosis Risk in Communities Study (ARIC), the Framingham Heart Study (FHS), the Women's Health Initiative (WHI) and a case-cohort study nested in the Prevención con Dieta Mediterránea Study (PREDIMED) (online supplemental table S1). The HCHS/SOL is a population-based cohort that recruited 16 415 Hispanic/Latino adults aged 18–74 years living in 4 US metropolitan areas.[15] A comprehensive battery of interviews and a clinical assessment with fasting blood draw were conducted at in-person clinic visits during 2008–2011 (baseline) and 2014–2017 (visit 2). Usual dietary intake was estimated using the National Cancer Institute methodology based on two 24 hours dietary recalls administered at baseline.[16] The ARIC study enrolled mostly white and black participants aged 45–64 years from four communities in the USA in 1987–1989.[17] The FHS was initiated in 1971 and we included FHS participants aged 40–65 years who attended the fifth examination (1991–1995).[18] The WHI study was launched in 1993 enrolling US women aged 50–79 years.[19] We also used data from a case-cohort study nested in the PREDIMED study which is a multicentre trial initiated in 2008.[20 21]

An expanded description of study populations, data collection and statistical analyses is provided in online supplemental file 1. All participants gave written informed consent.

Patient and public involvement

Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

Ascertainment of incident T2D

In all studies, participants free of diabetes at baseline who met at least one of the following criteria during the follow-up visits or telephone interviews were defined as incident T2D cases: fasting time >8 hours and fasting glucose ≥7.0 mmol/L (126 mg/dL), fasting ≤8 hours and non-fasting glucose ≥11.1 mmol/L (200 mg/dL), 2-hour post-oral glucose tolerance test glucose ≥11.1 mmol/L (200 mg/dL), homeglobine A1c ≥6.5%, treatment with antidiabetic medications or self-reported physician-diagnosed diabetes.

Metabolomic profiling

In HCHS/SOL, serum metabolomic profiling was performed using the discoveryHD4 platform at Metabolon (Durham, North Carolina, USA) in 3972 participants randomly selected from the whole cohort at baseline.[22] Eleven tryptophan metabolites, including tryptophan, serotonin, five kynurenine-pathway metabolites (kynurenine, kynurenate, xanthurenate, quinolinate and picolinate), and four indole derivatives (indoleacetate, indolelactate, indolepropionate and indoxyl sulfate) (online supplemental figure S1), were captured by an untargeted liquid chromatography-mass spectrometry (LC-MS) approach. In ARIC, seven tryptophan metabolites were available in the baseline serum metabolomics data measured by a similar LC-MS approach at Metabolon Inc.[17] In other studies, baseline plasma tryptophan metabolites (eight in FHS; five in PREDIMED; seven in WHI) were measured using LC-MS approaches at the Broad Institute (Cambridge, Massachusetts, USA).[18 19 21] Metabolomic approaches at both Metabolon and the Broad Institute are semiquantitative. We performed inverse normal transformation on relative levels of metabolites and conducted analyses separately within each study.

Genome-wide genotyping and imputation

Genotyping was performed using a customised Illumina array (15 041 502 B3; llmina Omni 2.5M array plus ~150K custom SNPs) in HCHS/SOL,[23] the Affymetrix V.6.0 chip in ARIC,[24] and the Affymetrix 500K and a 50K Human Gene Focused Panel in FHS.[9] Genome-wide imputation was carried out based on the 1000 Genomes Project phase 3 reference panel in HCHS/SOL and ARIC, and the HapMap CEU population reference panel in FHS.

Metagenomic sequencing and taxonomic profiling

Metagenomics sequencing was performed on DNA extracted from faecal samples collected by Flinders Technology Association (FTA) card from 3035 HCHS/SOL participants enrolled in a gut microbiome ancillary study during the HCHS/SOL Visit 2,[25] by a novel shallow-coverage method of shotgun sequencing using Illumina platforms.[26] To account for variability in sequencing depth, centred log-ratio transformation was applied to taxonomic abundances using R/microbiome. Ninety-two bacterial genera with average relative abundance ≥0.01% were included in the current analyses.

Statistical analysis

Figure 1 shows a workflow of our analysis. In stage I, we examined associations of circulating tryptophan metabolites with incident T2D, host genetics, dietary intake and gut microbiota using data from multiple studies (online supplemental table S1). Cox regression was used to estimate HRs and 95% CIs of incident T2D per SD increment in metabolites in each cohort separately, adjusting for demographic, social, behavioural and health-related factors, and other study-specific covariates (online supplemental table S2). Results from each of the cohorts were combined using a fixed-effect meta-analysis. GWAS of standardised metabolite levels were conducted separately in 3933 HCHS/SOL participants, 1509 ARIC white participants and 1772 ARIC black participants, controlling for age, sex, population stratification and other study-specific covariates. GWAS summary statistics for metabolites in 1438 whites from FHS, were obtained from a

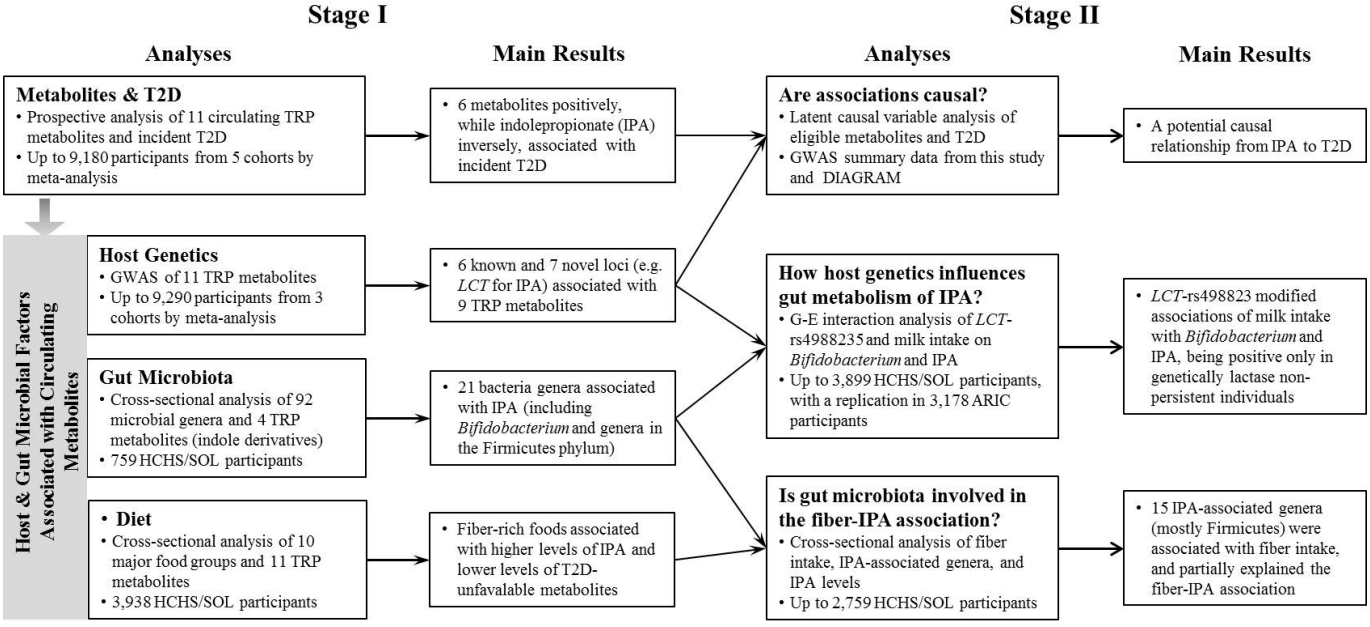

Figure 1 Overview of the workflow integrating host genetics, diet, gut microbiota and circulating metabolites in relation to T2D. Eleven tryptophan (TRP) metabolites included TRP, serotonin, five kynurenine-pathway metabolites (kynurenine, kynurenate, xanthurenate, quinolinate and picolinate), and four indole derivatives (indoleacetate, indolelactate, indolepropionate (IPA) and indoxyl sulfate). *LCT*-rs498823, a function variant related to lactase persistence. ARIC, Atherosclerosis Risk in Communities Study; DIAGRAM, Diabetes Genetics Replication and Meta-analysis; GWAS, genome-wide association study; HCHS/SOL, Hispanic community health Study/Study of Latinos; IPA, indolepropionic acid; T2D, type 2 diabetes.

previous publication.[9] Meta-analyses of GWAS summary statistics were conducted using METAL.[27] Associations of serum metabolites with 10 food groups which capture commonly consumed foods, three macronutrients, and two nutrients of interest (fibre and tryptophan) were analysed using multivariable linear regression in 3938 HCHS/SOL participants. A medication analysis using multiple mediator models[28] was performed to examine the potential mediating effect of serum tryptophan metabolites on the association between the overall diet quality, measured by the Alternate Healthy Eating Index 2010 (AHEI-2010),[29] and incident T2D in 2,821 HCHS/SOL participants. Associations of 92 gut bacterial genera with four indole derivatives were assessed by multivariable linear regression in 759 HCHS/SOL participants. Based on findings from stage I analyses, we performed multiple explanatory analyses in stage II (figure 1 and online supplemental table S3). We used linkage disequilibrium score regression[30] to estimate genetic heritability of metabolites and their genetic correlations with T2D. We applied the latent causal variable (LCV) model, which has been recommended to distinguish genetic correlation from causation, to test potential causal relationships between metabolites and T2D, as conventional Mendelian Randomisation approaches might be confounded by genetic correlations reflecting shared aetiology.[31] GWAS summary statistics for metabolites in this study (up to 9290 participants) and those for T2D obtained from the Diabetes Genetics Replication and Meta-analysis (DIAGRAM) consortium (55 005 T2D cases and 400 308 controls),[32] were used in the genetic correlation analysis and the LCV models. In HCHS/SOL, we used multivariable linear regression to examine associations between fibre intake and indolepropionate-associated bacterial genera (n=2759), and compared associations between fibre intake and indolepropionate with and without adjustment for indolepropionate-associated bacterial genera (n=752). A mediation analysis using structural equation modelling[33] was conducted to examine whether indolepropionate-associated bacterial genera may partially explain the association between fibre intake and indolepropionate. In HCHS/SOL, we applied multivariable linear regression to examine associations of *LCT*-rs4988235 with milk intake (n=12 531), gut *Bifidobacterium* abundance (n=2368) and serum indolepropionate (n=3933). Multivariable linear regression was used to examine associations of milk intake with gut *Bifidobacterium* abundance and serum indolepropionate levels stratified by the *LCT*-rs4988235 genotype (lactase persistence AA/AG vs lactase non-persistence GG), and the interaction between *LCT*-rs4988235 and milk intake was examined by introducing an interaction term. To validate the interaction between milk intake and *LCT*-rs4988235 on indolepropionate, a replication analysis was performed in ARIC (1504 whites and 1674 blacks).

Analyses were performed using R software unless otherwise stated. In GWAS, $p<4.5\times10^{-9}$ (5.0×10^{-8}/11 metabolites) was considered as genome-wide significant, and a false discovery rate (FDR) <0.05 was considered as statistically significant for other primary analyses.

RESULTS

Tryptophan metabolites and incident T2D

Baseline characteristics of study participants are shown in online supplemental table S4. Among 2821 US Hispanics/Latinos without diabetes at baseline from HCHS/SOL, 367 incident T2D cases were identified during a median of 5.7 years of follow-up. Among 6359 participants, free of diabetes at baseline, with diverse racial/ethnic backgrounds from ARIC, FHS, WHI and PREDIMED, 1665 incident T2D cases were identified during follow-up. Of 11 metabolites, tryptophan, four kynurenine metabolites (kynurenine, kynurenate, xanthurenate and quinolinate) and indolelactate were positively associated with incident T2D, while indolepropionate was inversely associated with incident T2D after multivariable adjustment in combined analysis

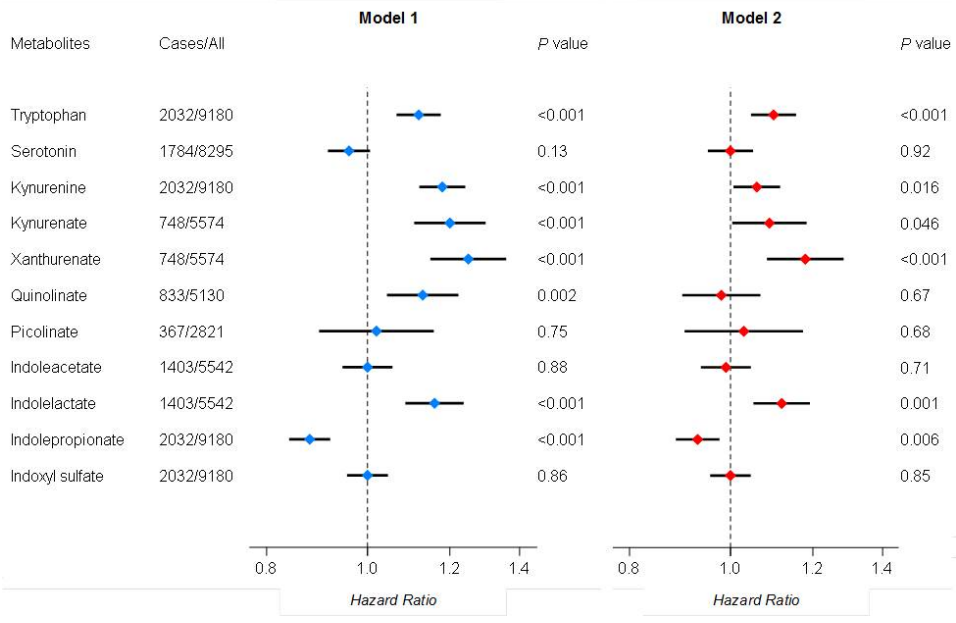

Figure 2 Associations between circulating tryptophan metabolite levels and incident type 2 diabetes. Data are HRs and 95% CI of incident type two diabetes per SD increment in metabolite levels, adjusted for age, sex, smoking, alcohol consumption, education, family income, family history of diabetes, self-reported hypertension and/or antihypertensive medication use, self-reported dyslipidaemia and/or lipid-lowering medication use and other study-specific covariates (model1); and further adjusted for body mass index and waist-to-hip ratio (model 2). Results across five studies were combined by fixed-effect meta-analysis.

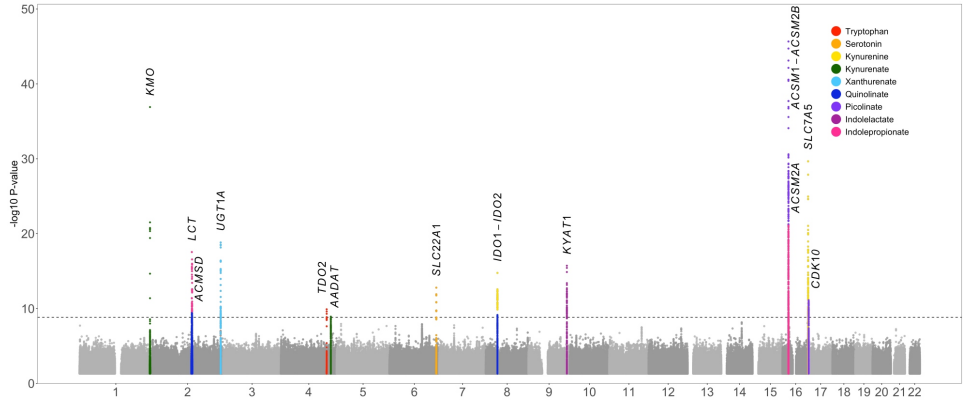

Figure 3 Manhattan plot for GWAS of circulating tryptophan metabolite levels. Meta-analyses of GWAS in up to 9290 individuals from HCHS/SOL, ARIC, and FHS identified 13 loci for nine tryptophan metabolites (colour indicated in inset). The significant p value threshold is 4.5×10–9 (indicated by a dash line). ARIC, Atherosclerosis Risk in Communities Study; FHS, Framingham Heart Study; GWAS, Genome-wide association studies; HCHS/SOL, Hispanic Community Health Study/Study of Latinos.

of all studies (all FDR <0.05; figure 2, model 1). Results were generally consistent across HCHS/SOL and the other four studies (online supplemental table S5). The observed associations were attenuated but remained significant after further adjusting for obesity measures including body mass index (BMI) and waist-to-hip ratio, except for quinolinate (figure 2, model 2). Further adjustment for blood lipids, blood pressure or physical activity and dietary quality did not materially change these associations (online supplemental table S5).

Among 2821 HCHS/SOL participants without diabetes at baseline, metabolites that were positively associated with T2D (ie, tryptophan, kynurenine, kynurenate, xanthurenate, quinolinate and indolelactate) showed weak-to-moderate correlations with each other (Spearman's r=0.11 to 0.63) (online supplemental figure S2), and positive correlations with multiple cardiometabolic traits, especially fasting insulin, Homeostatic Model Assessment for Insulin Resistance (HOMA-IR) and BMI (online supplemental figure S3). Indolepropionate, the only metabolite inversely associated with T2D, was not correlated with other metabolites (Spearman's r=-0.05 to 0.06), and showed significant, although weak, inverse correlations with BMI and a few other cardiometabolic traits.

Host genetics and tryptophan metabolites

Our genome-wide meta-analyses (n=up to 9290) identified 21 independent signals at 13 loci associated with nine of 11 tryptophan metabolites ($p<4.5\times10^{-9}$) (figure 3 and online supplemental table S6). Genetic variants at seven loci have not been previously associated with the corresponding metabolites, including those in or near *SLC22A, IDO1-IDO2, AADAT*, *ACMSD, ACSM2B-ACSM1*, *CDK10* and *LCT*. We confirmed known genetic associations at six loci.[9–11] When the threshold of significance was relaxed to traditional genome-wide significance ($p<5.0\times10^{-8}$), we found 16 additional loci associated with tryptophan metabolites (online supplemental table S6). Many newly identified and confirmed signals reside in genomic regions harbouring genes involved in host kynurenine pathway metabolism (eg, *TDO2*, *IDO1-IDO2, KMO*, *AADAT* and *ACMSD*) or transportation of tryptophan metabolites (eg, *SLC7A5*, *SLC22A1* and *SLC16A10*).

Based on GWAS summary statistics from our meta-analysis, genome-wide SNP-based heritability (h^2) was estimated at 13.0% (SE=4.9%) for serotonin, 10.7% (5.8%) for indolepropionate, 7.4% (4.8%) for kynurenine and 0%–7.0% for other metabolites (online supplemental table S7). As expected, these genome-wide SNP-based heritability estimates were much lower than those estimated using the classical twin model and were generally higher than those estimated based on a few genome-wide significant variants in previous studies (online supplemental table S7).[9–11] We then examined potential causal relationships between three metabolites (serotonin, indolepropionate and kynurenine), which had heritability estimates meeting the criteria for LCV models,[31] and T2D using GWAS summary statistics for metabolites in our study (n=up to 9290) and those for T2D obtained in the DIAGRAM (55 005 T2D cases and 400 308 controls).[32] Indolepropionate showed a potential causal relationship with T2D (genetic causality proportion=76%, $p=1.6\times10^{-24}$) (online supplemental table S7).

Dietary intake and tryptophan metabolites

In 3938 HCHS/SOL participants, we observed significant associations of higher intakes of vegetables, fruits, whole grains, nuts and legumes, and lower intakes of refined grains and red meat, with higher serum indolepropionate levels (figure 4A). Intakes of some fibre-rich foods which were positively associated with indolepropionate showed inverse associations with other indole derivatives and most kynurenine-pathway metabolites. Mutual adjustment for other food groups did not materially change the results (online supplemental table S8). Consistently, higher fibre intake was associated with higher indolepropionate ($p=7.3\times10^{-60}$), and with lower levels of other indole derivatives and most kynurenine-pathway metabolites (figure 4B). These associations were independent of intakes of macronutrients and tryptophan (online supplemental table S9). Intakes of some protein-rich foods (eg, red meat, poultry and dairy) and tryptophan were positively associated with serum levels of tryptophan, most kynurenine-pathway metabolites, and indoxylsulfate (figure 4A and online supplemental tables S8 and S9). Our mediation analysis in 2821 HCHS/SOL participants without diabetes at baseline indicated a significant meditating effect of these tryptophan metabolites on the association between the overall diet quality (ie, AHEI-2010) and incident T2D (proportion mediated=61.5%; p=0.01).

Gut microbiota and indole derivatives

As indole pathway is carried out mostly by gut microbiota,[4] we examined associations between 92 gut microbial genera and serum levels of four indole derivatives in 759 HCHS/SOL participants. We focused on indolepropionate and indolelactate as these two indole derivatives were significantly associated with

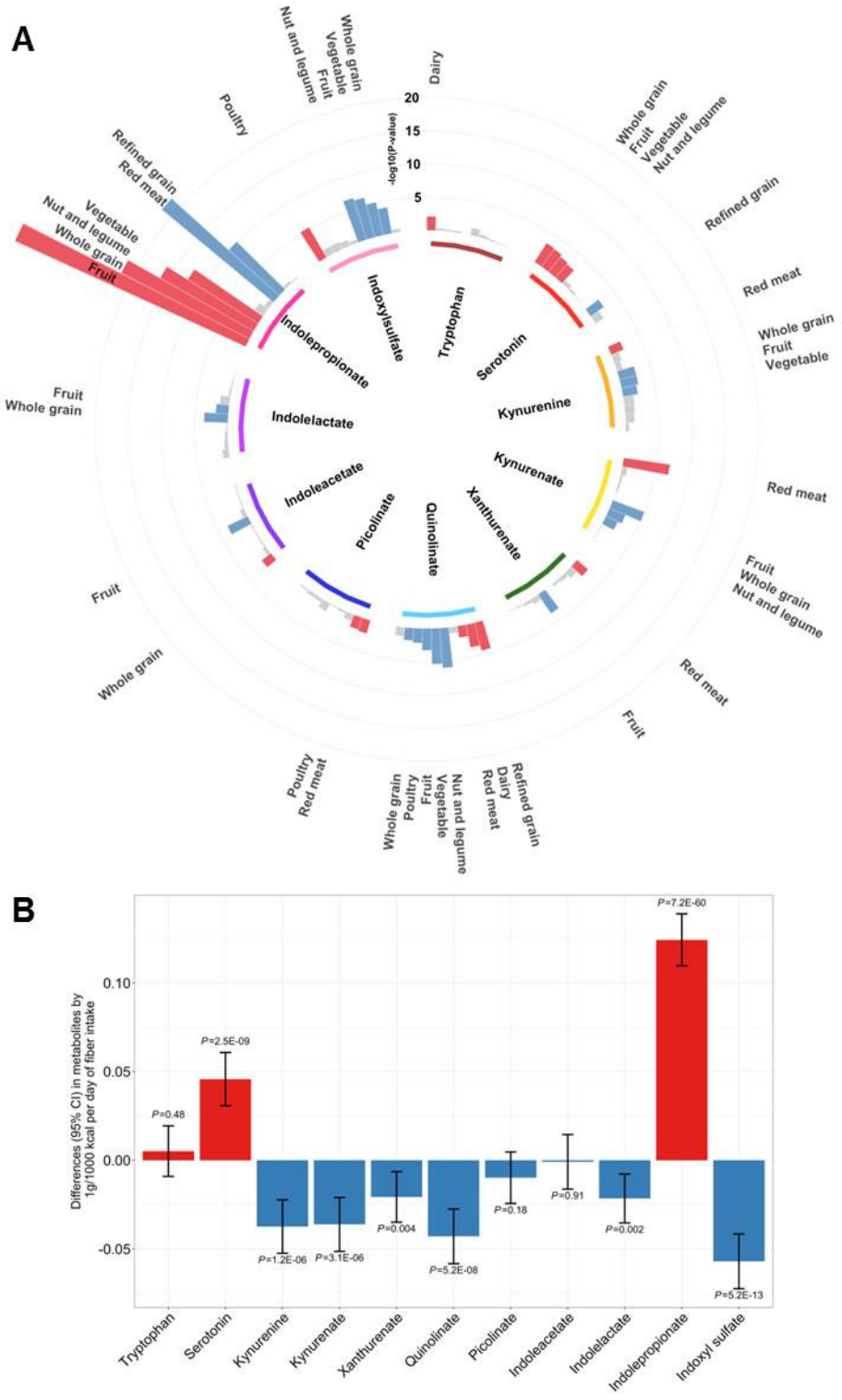

Figure 4 Dietary intake and serum tryptophan metabolite levels. (A) Polar plot for associations of 10 major food groups with serum tryptophan metabolites in the HCHS/SOL. Red: positive associations (FDR <0.05); blue, inverse associations (FDR <0.05). (B) Differences (95% CI) in serum tryptophan metabolite levels (inverse normal transformed) associated with 1 g/1000 Kcal per day of dietary fibre intake in the HCHS/SOL. FDR, false discovery rate; HCHS/SOL, Hispanic Community Health Study/Study of Latinos.

incident T2D in our study, and identified 21 genera significantly associated with indolepropionate (FDR <0.05) but none associated with indolelactate (online supplemental table S10). In addition, five bacterial genera were significantly associated with indoleacetate and 11 were associated with indoxyl sulfate.

The 21 indolepropionate-associated genera span 3 phyla (Firmicutes, n=16; Actinobacteria, n=3; and Bacteroidetes, n=2) (figure 5A). When we included all 21 genera in the linear regression model on indolepropionate simultaneously, associations for these genera (especially those in Firmicutes) were

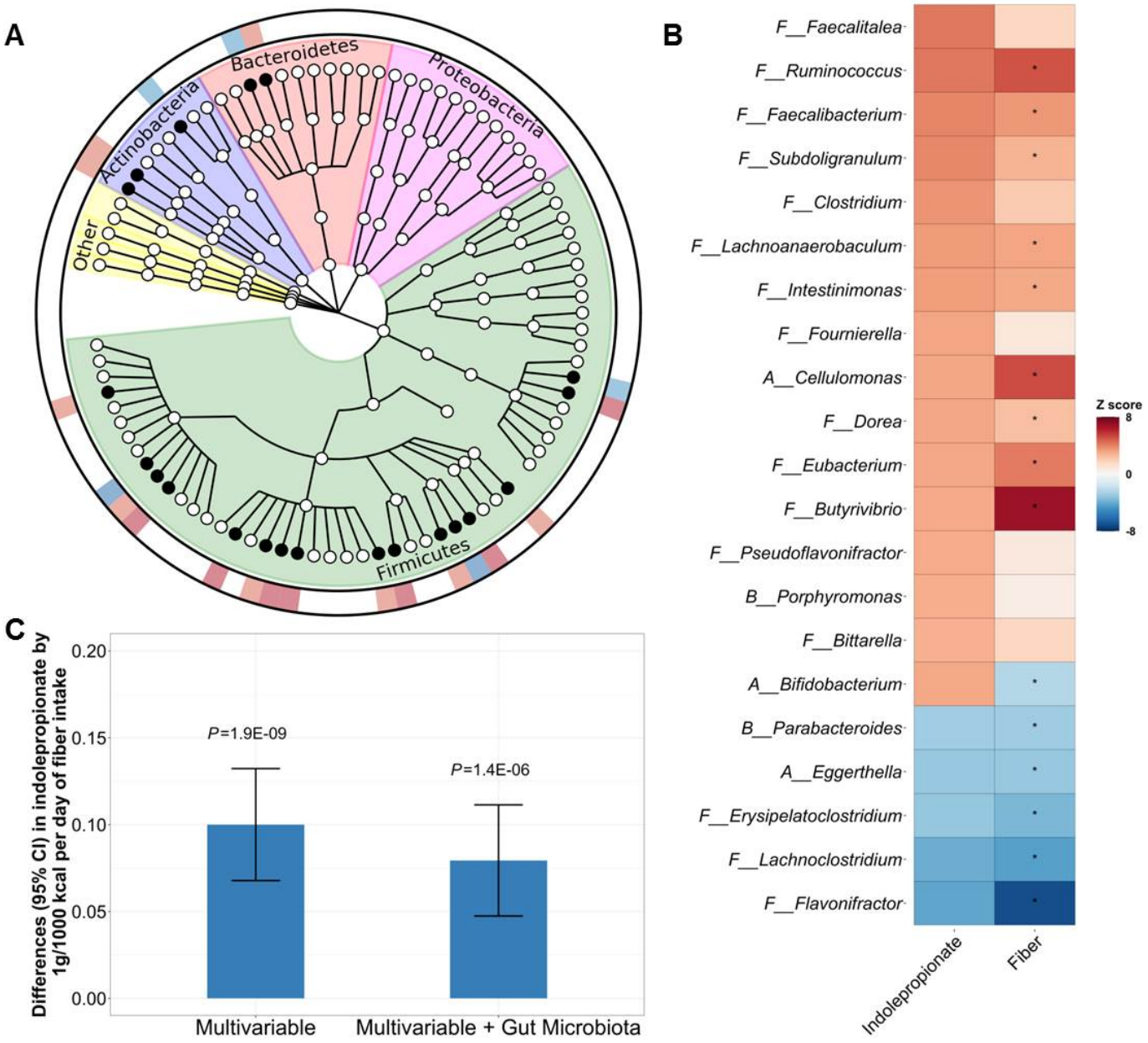

Figure 5 Dietary fibre intake, gut microbiota and serum indolepropionate. (A) Phylogenetic tree of taxonomic features in association with host serum indolepropionate levels in the HCHS/SOL. A total of 21 gut microbial genera significantly associated with serum indolepropionate (FDR <0.05) are indicated by solid circles. Data showing in the outer ring are effect sizes (positive, red; inverse, blue) of gut microbiota genera on serum indolepropionate. (B) Associations of 21 indolepropionate-assocaited gut microbial genera with dietary fibre intake in the HCHS/SOL. To show comparable estimates for the associations of gut microbial genera with indolepropionate and fibre intake, data are presented as Z-scores (regression coefficients/SEs). *FDR <0.05 for the associations between dietary fibre intake and gut microbial genera. (C) Associations between dietary fibre intake and serum indolepropionate levels with and without adjustment for gut microbiota (20 indolepropionate-associated gut microbial genera) in the HCHS/SOL. Bifidobacterium, which showed opposite associations with indolepropionate and fibre intake, was not included. FDR, false discovery rate; HCHS/SOL, Hispanic Community Health Study/Study of Latinos.

greatly attenuated or abolished, while the association between *Bifidobacterium* and indolepropionate did not change (online supplemental figure S4).

Fibre intake, gut microbiota and indolepropionate

In 2759 HCHS/SOL participants with diet and gut microbiome data, all indolepropionate-associated bacterial genera were associated with fibre intake (15 genera showing FDR <0.05) with the same directions as those associations between bacterial genera and indolepropionate, except for *Bifidobacterium* (figure 5B). In 752 HCHS/SOL participants with diet, metabolomics and gut microbiome data, the association between fibre intake and indolepropionate was attenuated after further adjustment for the 20 indolepropionate-associated bacterial genera excluding *B*ifidobacterium (figure 5C). The attenuation was similar when including *Bifidobacterium* in the model. We also found a potential mediating effect of these 20 indolepropionate-associated bacterial genera on the association between fibre intake and indolepropionate (proportion mediated=22.3%; p=0.003). These results suggested that these 20 indolepropionate-associated

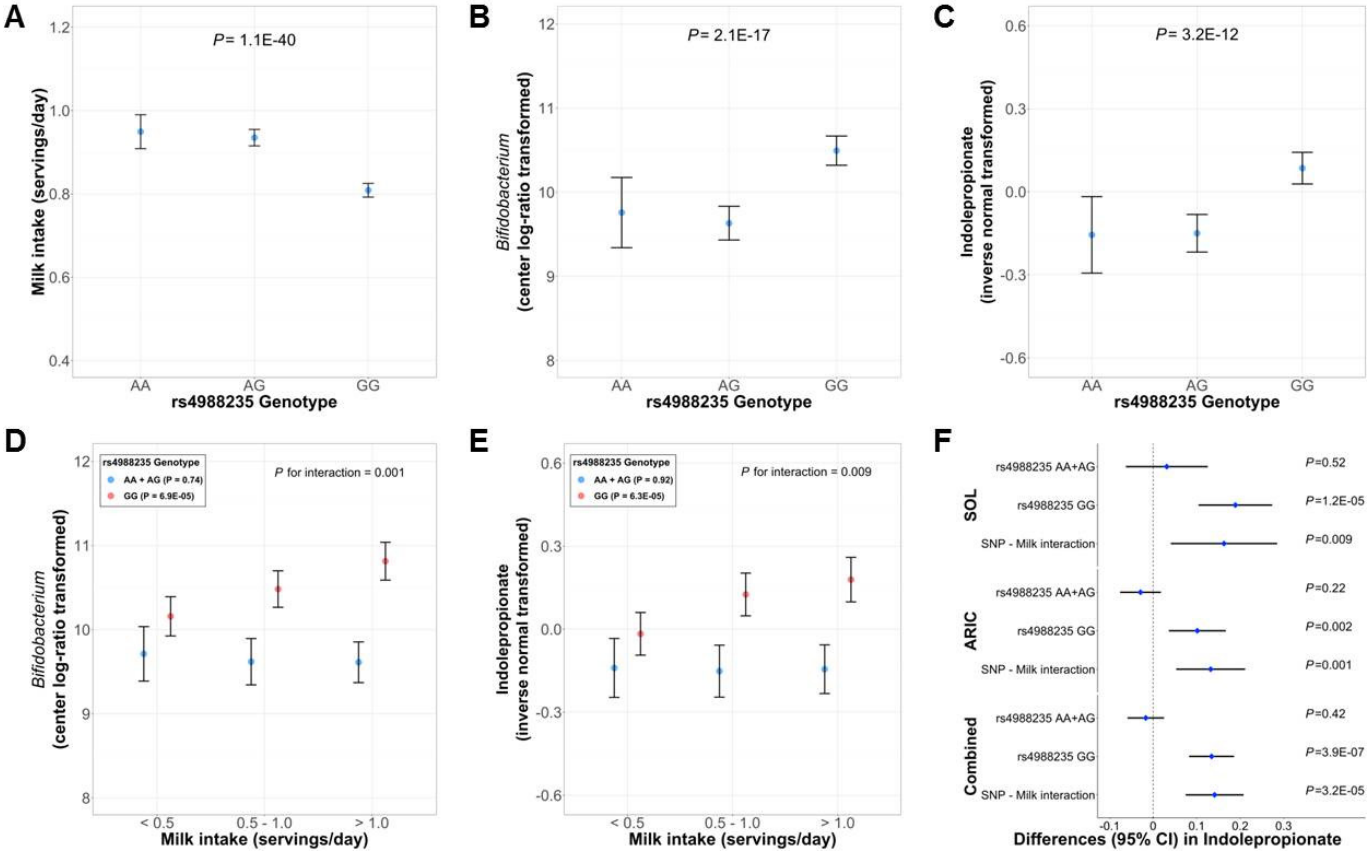

Figure 6 Host LCT genotype, milk intake, gut Bifidobacterium and serum indolepropionate (A) adjusted means and 95% CIs of milk intake (servings/day) according to LCT-rs49883235 genotypes in the HCHS/SOL. (B) Adjusted means and 95% CI of gut Bifidobacterium abundance (centre log-ratio transformed) according to LCT-rs49883235 genotypes in the HCHS/SOL. (C) Adjusted means and 95% CIs of serum indolepropnate levels (inverse normal transformed) according to LCT-rs49883235 genotypes in the HCHS/SOL. (D) Adjusted means and 95% CI of gut Bifidobacterium abundance (centre log-ratio transformed) according to milk intake stratified by the LCT-rs49883235 genotype in the HCHS/SOL. (E) Adjusted means and 95% CI of serum indolepropnate levels (inverse normal transformed) according to milk intake stratified by the LCT-rs49883235 genotype in the HCHS/SOL. (F) Differences and 95% CI in serum indolepropnate levels (inverse normal transformed) associated with one serving per day of milk intake according to the LCT-rs49883235 genotype in the HCHS/SOL and ARIC separately, and combined by meta-analysis. ARIC, Atherosclerosis Risk in Communities Study; HCHS/SOL, Hispanic Community Health Study/Study of Latinos.

bacterial genera may partially explain the association between fibre intake and indolepropionate, while *Bifidobacterium* may be involved in other pathways related to indolepropionate.

Host *LCT*, gut *Bifidobacterium* and indolepropionate

We then focused on gut *Bifidobacterium* in association with indolepopionate, as gut *Bifidobacterium* abundance has been related to a host functional *LCT* variant (rs4988235)[34 35] and our GWAS also identified *LCT* as a novel locus for indolepropionate. *LCT*-rs4988235 is a known variant which determines lactase persistence in adulthood (AA/AG is related to lactase persistence and GG is related to lactase non-persistence).[36] In line with previous evidence,[34–36] the rs4988235-G allele was associated with lower milk intake ($p=1.1\times10^{-40}$; n=12 531) (figure 6A) and higher gut *Bifidobacterium* abundance ($p=2.1\times10^{-17}$; n=2368) (figure 6B) in HCHS/SOL. In our GWAS, rs4988235-G allele was associated with higher circulating indolepropionate levels ($p=3.2\times10^{-17}$ in meta-analysis, n=9290; p for heterogeneity=0.51) ($p=3.2\times10^{-12}$ in HCHS/SOL, n=3933; figure 6C). When we included both *LCT*-rs4988235 and *Bifidobacterium* in the multivariable linear regression model on indolepropionate (n=752), *Bifidobacterium*, but not *LCT*-rs4988235, was significantly associated with indolepropionate.

Consistent with prior evidence,[34 35] we found that milk intake was positively associated with gut *Bifidobacterium* abundance only among lactase non-persistent participants (rs4988235 GG, $p=6.9\times10^{-5}$) but not among those with lactase persistence (rs4988235 AG+GG; p=0.74) in HCHS/SOL (*p-interaction*=0.001; n=2342) (figure 6D). Paralleling the *LCT*-milk interaction on gut *Bifidobacterium*, we identified a novel interaction between milk intake and *LCT* genotype on serum indolepropionate (p interaction=0.009; n=3899). Milk intake was positively associated with serum indolepropionate levels only among lactase non-persistent individuals ($p=6.3\times10^{-5}$) but not in those with lactase persistence (p=0.92) (figure 6E). This significant interaction was replicated in ARIC (p interaction=0.001; n=3178) (figure 6F). *LCT*-rs4988235 did not show significant associations with intakes of dairy products low in lactose (eg, yoghurt, cheese) or significant interactions with other dairy products on gut *Bifidobacterium* abundance or serum indolepropionate levels (data not shown).

DISCUSSION

In large-scale populations with diverse racial/ethnic backgrounds, our study demonstrated that circulating levels of kynurenine-pathway metabolites, a group of host tryptophan

catabolites, including kynurenine, quinolinate, kynurenate and xanthurenate[7] were associated with increased risk of T2D. We also found that higher intakes of animal-based, protein-rich foods and lower intakes of plant-based, fibre-rich foods were associated with higher circulating levels of kynurenine-pathway metabolites, but the associations between kynurenine-pathway metabolites and T2D did not change after further adjustment for diet quality score. This suggests that these metabolites could be potential mediators linking unhealthy diets with increased risk of T2D rather than simple biomarkers reflecting adverse dietary effects. Moreover, these kynurenine-pathway metabolites were positively correlated with obesity measures and insulin resistance, and obesity may partially explain our observed associations between these metabolites and T2D. These findings are in line with previous evidence and support the notion that activation of the kynurenine pathway by obesity and related inflammation may affect insulin signalling and contribute to increased risk of T2D.[2 3 7 21]

Indole derivatives, a group of microbial tryptophan catabolites, are generally beneficial for human health.[4] Higher circulating indolepropionate has been associated with lower risk of T2D,[8 12] but it was argued that this association might just reflect beneficial effects of dietary fibre intake on T2D.[14] Our study documented the beneficial association between indolepropionate and T2D and further suggested potential causality. This is consistent with the potential role of indolepropionate in antioxidation, anti-inflammation and amelioration of glucose metabolism.[4]

As little evidence suggests that indolepropionate can be derived from fibre fermentation, the strong positive association between fibre intake and circulating indolepropionate is intriguing,[8 12 14] but may be explained, in part, by a potential novel pathway suggested by our integrative analysis. Tryptophan is the sole source for indolepropionate production which is suggested as completely gut microbiota-dependent in mice,[37] involving bacterial species mostly in the *Clostridium* genus.[38] Consistently, a majority of identified gut bacterial genera in our study, including *Clostridium*,[38] showed positive associations with indolepropionate. Most of these genera are members of Firmicutes, a phylum that includes many species use dietary fibre as main energy source.[39] Catabolism of aromatic amino acids including tryptophan has been demonstrated in Firmicutes but not in other phyla.[38] We also found several indolepropionate-bacterial genera in other phyla which might be related to fibre intake, although it is unknown whether they are involved in the indolepropionate production. For example, *Cellulomonas*, a genus in Actinobacteria, is known to degrade cellulose,[40] a type of fibre found in plant cell walls. These findings suggest that higher fibre intake may increase populations of fibre-using bacteria,[39] some of which may have the capability to produce indolepropionate or its substrates from tryptophan,[4] thus shifting host tryptophan-to-kynurenine catabolism more towards gut microbial indolepropionate production. However, it should be noted that the association between fibre intake and indolepropionate was not fully explained by the identified bacteria in our study. Gut bacteria involved in this pathway might not be fully captured by our faecal metagenomics. A notable limitation of our study is that the assessments of diet and serum metabolites preceded faecal sample collection by a median of 7 years. Although the human gut microbiome was found to be notably stable over a long period,[41] the 7-year time lag might attenuate the associations of the gut microbiota with diet and metabolites in this study. It is possible that we would observe stronger associations of gut microbiota with fibre intake and serum indolepropionate with concurrently collected data. Nevertheless, our findings suggest indolepropionate production, in addition to short-chain fatty acid production,[39] as a potential novel microbial metabolite pathway for beneficial effects of dietary fibre on human cardiometabolic health.

Another novel finding of this study is that a lactase persistence-determining variant at *LCT* was associated with circulating indolepropionate, through an apparent interaction with milk intake. This might be related to an indolepropionate-associated gut bacterium identified in this study, *Bifidobacterium*, which has been associated with host *LCT* and milk intake.[34 35] Compared with lactase persistent individuals, lactase non-persistent individuals cannot hydrolyze lactose after consuming milk and thus have more lactose in the gut as an energy source for *Bifidobacterium* growth,[34 35] which may then contribute to higher indolepropionate production. Indeed, although it is unknown whether *Bifidobacterium* has the capability to produce indolepropionate, many strains in the *Bifidobacterium* genus have been found to produce indolelactate,[42 43] a substrate for indolepropionate. Moreover, both human[44–47] and animal studies[48] suggested a potential protective role of gut *Bifidobacterium* in T2D. Taken together, our observations extend the previously identified host gene–diet interaction on gut microbiota[34 35] to microbiota-produced metabolites in host circulation, and suggest microbial indole derivative production as a potential mechanism through which gut *Bifidobacterium* is associated with T2D. Due to limitations of shallow shotgun sequencing data,[26] we did not examine *Bifidobacterium* species or strains, or functional features for indole derivative production which need to be clarified in future studies.

The other two indole derivatives, indoleacetate and indolelactate, have been shown to act through aryl hydrocarbon receptor activation,[4] which could reduce inflammation and insulin resistance.[5] However, we did not find beneficial associations of these two metabolites with T2D. In contrast, indolelactate was associated with increased risk of T2D in our study, and inconsistent associations between indolelactate and insulin resistance were also reported in previous studies.[49 50] Interestingly, we found that serum indolelactate was more closely correlated with kynurenine-pathway metabolites than other indole derivatives, and host factors (eg, genetic variants in *KYAT1*,[10] a gene involved in host tryptophan-kynurenine metabolism) rather than gut microbial factors were associated with circulating indolelactate levels. Further studies are warranted to clarify the relationship between circulating and faecal indole derivatives and their associations with T2D.

In summary, circulating tryptophan, several kynurenine-pathway metabolites and indolelactate showed adverse associations with incident T2D, while indolepropionate showed a beneficial association with incident T2D. We identified multiple host genetic, dietary and gut microbial factors associated with these metabolites. In particular, higher fibre intake and milk intake (only among genetically lactase non-persistent individuals) were associated with higher circulating levels of indolepropionate possibly through the host–microbial cross-talk shifting tryptophan metabolism towards gut microbial indolepropionate production. It should be noted that our study is unable to make causal inference due to its observational nature, although our findings may have strong biological plausibility. These findings contribute to our understanding of the host–microbial cross-talk in tryptophan metabolism and its implications in human metabolic health and disease, and may help to identify high-risk individuals based on circulating metabolite profiles for targeted interventions through dietary intervention and gut microbiota modification.

Author affiliations
[1]Department of Epidemiology and Population Health, Yeshiva University Albert Einstein College of Medicine, Bronx, New York, USA
[2]Department of Nutrtion, Harvard University T H Chan School of Public Health, Boston, Massachusetts, USA
[3]Department of Epidemiology, Harvard University T H Chan School of Public Health, Boston, Massachusetts, USA
[4]Department of Epidemiology, Human Genetics, and Environmental Sciences, The University of Texas Health Science Center at Houston School of Public Health, Houston, Texas, USA
[5]Diabetes Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA
[6]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA
[7]Division of Women's Health, Department of Medicine at Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA
[8]Department of Preventive Medicine and Public Health, University of Navarra, Pamplona, Spain
[9]CIBER Fisiopatologıa de la Obesidad y Nutricion, Instituto de Salud Carlos III, Madrid, Spain
[10]Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA
[11]Departments of Pediatrics, Microbiology and Immunology, and Gynecology and Women's Health, Yeshiva University Albert Einstein College of Medicine, Bronx, New York, USA
[12]Division of General Internal Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA
[13]Department of Mathematics, University of Houston, Houston, Texas, USA
[14]Department of Internal Medicine, The University of Texas Southwestern Medical Center, Dallas, Texas, USA
[15]Programs in Metabolism and Medical & Population Genetics, Eli and Edythe L. Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA
[16]Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA
[17]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA
[18]Human Nutrition Unit, Faculty of Medicine and Health Sciences, Universidad Rovira i Virgili Departamento de Medicina y Cirurgía, Reus, Spain
[19]Department of Epidemiology, The University of Iowa College of Public Health, Iowa City, Iowa, USA
[20]Institute of Minority Health Research, University of Illinois College of Medicine, Chicago, Illinois, USA
[21]Department of Laboratory Medicine and Pathology, University of Minnesota Medical Center, Minneapolis, Minnesota, USA
[22]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[23]Division of Preventive Medicine, Department of Medicine at Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA
[24]Metabolomics Platform, Eli and Edythe L. Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA
[25]Channing Division of Network Medicine, Department of Medicine at Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA
[26]Department of Pediatrics, School of Medicine; Center for Microbiome Innovation, Department of Computer Science and Engineering, Jacobs School of Engineering, University of California San Diego, La Jolla, California, USA
[27]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA

Correction notice This article has been corrected since it published Online First. The equal contributor statement has been added.

Contributors QQ: study concept and design; study supervision; obtained funding; acquisition of data; analysis and interpretation of data; drafting of the manuscript; critical revision of the manuscript for important intellectual content. JL: study concept and design; acquisition of data; analysis and interpretation of data; drafting of the manuscript; critical revision of the manuscript for important intellectual content; BY: study concept and design; obtained funding; acquisition of data; analysis and interpretation of data; critical revision of the manuscript for important intellectual content; J-YM, JC, JM, JH, MR-C, CR, ZW, MU, G-CC, BP, WW, QN, EVF and TW: analysis and interpretation of data; critical revision of the manuscript for important intellectual content; MG, TW, RG, JD, JSS, WB, DP, MD, BT, JC, JM, MAMG, ES, CC and JM: acquisition of data; critical revision of the manuscript for important intellectual content. KMR, FH, RB, RB, EB and RK: obtained funding; acquisition of data; critical revision of the manuscript for important intellectual content.

Funding This work is supported by R01-DK119268 (QQ) from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), R01-MD011389 (RK, RDB and RCK) from the National Institute on Minority Health and Health Disparities, and R01-HL060712 (QQ and FBH) from the National Heart, Lung and Blood Institute (NHLBI). Other funding sources for this study include UM1-HG008898 from the National Human Genome Research Institute; R01-HL140976, and R01-HL136266 from the NHLBI; and R01-DK120870, R01 DK126698 and the New York Regional Center for Diabetes Translation Research (P30DK111022) from NIDDK. Support for metabolomics data was graciously provided by the JLH Foundation (Houston, Texas). JL is supported by 9-17-CMF-011 from the American Diabetes Association (ADA); K99-DK122128 from NIDDK; and a pilot and feasibility grant from the NIDDK-funded Boston Nutrition Obesity Research Center (P30-DK046200). BY is supported by the American Heart Association (17SDG33661228); and R01-HL141824 and R01-HL142003 from NHLBI. JM and JD are supported by U01-DK078616 from NIDDK. JM is supported by a postdoctoral fellowship funded by the European Commission Horizon 2020 programme; Marie Skłodowska-Curie actions (H2020-MSCA-IF-2015-703787); and P30DK040561 from NIDDK. Funding and acknowledgment information for each participating studies is provided in online supplements.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Study protocols were approved by the Institutional Review Boards at all participating institutions.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data are available on reasonable request. Anonymised individual participant data are available on reasonable request and with the agreement of the authors of original studies.Gut microbiome data in this study are deposited in QIITA, ID 11666 and EMBL-EBI ENA, ERP117287.

ORCID iDs
Qibin Qi http://orcid.org/0000-0002-2687-1758
Jun Li http://orcid.org/0000-0003-3519-8638
Jordi Merino http://orcid.org/0000-0001-8312-1438
Zheng Wang http://orcid.org/0000-0003-2052-5552

REFERENCES

1 Badawy AA-B. Kynurenine pathway of tryptophan metabolism: regulatory and functional aspects. *Int J Tryptophan Res* 2017;10:117864691769193.
2 Favennec M, Hennart B, Caiazzo R, *et al*. The kynurenine pathway is activated in human obesity and shifted toward kynurenine monooxygenase activation. *Obesity* 2015;23:2066–74.
3 Oxenkrug G. Insulin resistance and dysregulation of tryptophan-kynurenine and kynurenine-nicotinamide adenine dinucleotide metabolic pathways. *Mol Neurobiol* 2013;48:294–301.
4 Roager HM, Licht TR. Microbial tryptophan catabolites in health and disease. *Nat Commun* 2018;9:3294.
5 Laurans L, Venteclef N, Haddad Y, *et al*. Genetic deficiency of indoleamine 2,3-dioxygenase promotes gut microbiota-mediated metabolic health. *Nat Med* 2018;24:1113–20.
6 Wang TJ, Larson MG, Vasan RS, *et al*. Metabolite profiles and the risk of developing diabetes. *Nat Med* 2011;17:448–53.
7 Vangipurapu J, Fernandes Silva L, Kuulasmaa T, *et al*. Microbiota-Related metabolites and the risk of type 2 diabetes. *Diabetes Care* 2020;43:1319–25.
8 de Mello VD, Paananen J, Lindström J, *et al*. Indolepropionic acid and novel lipid metabolites are associated with a lower risk of type 2 diabetes in the Finnish diabetes prevention study. *Sci Rep* 2017;7:46337.
9 Rhee EP, Ho JE, Chen M-H, *et al*. A genome-wide association study of the human metabolome in a community-based cohort. *Cell Metab* 2013;18:130–43.
10 Shin S-Y, Fauman EB, Petersen A-K, *et al*. An atlas of genetic influences on human blood metabolites. *Nat Genet* 2014;46:543–50.
11 Long T, Hicks M, Yu H-C, *et al*. Whole-Genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat Genet* 2017;49:568–78.
12 Tuomainen M, Lindström J, Lehtonen M, *et al*. Associations of serum indolepropionic acid, a gut microbiota metabolite, with type 2 diabetes and low-grade inflammation in high-risk individuals. *Nutr Diabetes* 2018;8:35.
13 Pallister T, Jennings A, Mohney RP, *et al*. Characterizing blood metabolomics profiles associated with self-reported food intakes in female twins. *PLoS One* 2016;11:e0158568.

14 Menni C, Hernandez MM, Vital M, *et al*. Circulating levels of the anti-oxidant indoleproprionic acid are associated with higher gut microbiome diversity. *Gut Microbes* 2019;10:688–95.
15 Sorlie PD, Avilés-Santa LM, Wassertheil-Smoller S, *et al*. Design and implementation of the Hispanic community health Study/Study of Latinos. *Ann Epidemiol* 2010;20:629–41.
16 Siega-Riz AM, Sotres-Alvarez D, Ayala GX, *et al*. Food-group and nutrient-density intakes by Hispanic and Latino backgrounds in the Hispanic community health Study/Study of Latinos. *Am J Clin Nutr* 2014;99:1487–98.
17 Rebholz CM, Yu B, Zheng Z, *et al*. Serum metabolomic profile of incident diabetes. *Diabetologia* 2018;61:1046–54.
18 Merino J, Leong A, Liu C-T, *et al*. Metabolomics insights into early type 2 diabetes pathogenesis and detection in individuals with normal fasting glucose. *Diabetologia* 2018;61:1315–24.
19 Paynter NP, Balasubramanian R, Giulianini F, *et al*. Metabolic predictors of incident coronary heart disease in women. *Circulation* 2018;137:841–53.
20 Estruch R, Ros E, Salas-Salvadó J, *et al*. Primary prevention of cardiovascular disease with a Mediterranean diet supplemented with extra-virgin olive oil or nuts. *N Engl J Med Overseas Ed* 2018;378:e34.
21 Yu E, Papandreou C, Ruiz-Canela M, *et al*. Association of tryptophan metabolites with incident type 2 diabetes in the PREDIMED trial: a case-cohort study. *Clin Chem* 2018;64:1211–20.
22 Chen G-C, Chai JC, Yu B, *et al*. Serum sphingolipids and incident diabetes in a US population with high diabetes burden: the Hispanic community health Study/Study of Latinos (HCHS/SOL). *Am J Clin Nutr* 2020;112:57–65.
23 Conomos MP, Laurie CA, Stilp AM, *et al*. Genetic diversity and association studies in US Hispanic/Latino populations: applications in the Hispanic community health Study/Study of Latinos. *Am J Hum Genet* 2016;98:165–84.
24 Psaty BM, O'Donnell CJ, Gudnason V, *et al*. Cohorts for heart and aging research in genomic epidemiology (charge) Consortium: design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet* 2009;2:73–80.
25 Kaplan RC, Wang Z, Usyk M, *et al*. Gut microbiome composition in the Hispanic community health Study/Study of Latinos is shaped by geographic relocation, environmental factors, and obesity. *Genome Biol* 2019;20:219.
26 Hillmann B, Al-Ghalith GA, Shields-Cutler RR, *et al*. Evaluating the information content of shallow shotgun metagenomics. *mSystems* 2018;3:e00069–18.
27 Willer CJ, Li Y, Abecasis GR. Metal: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010;26:2190–1.
28 Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav Res Methods* 2008;40:879–91.
29 Chiuve SE, Fung TT, Rimm EB, *et al*. Alternative dietary indices both strongly predict risk of chronic disease. *J Nutr* 2012;142:1009–18.
30 Bulik-Sullivan BK, Loh P-R, Finucane HK, *et al*. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015;47:291–5.
31 O'Connor LJ, Price AL. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat Genet* 2018;50:1728–34.
32 Mahajan A, Taliun D, Thurner M, *et al*. Fine-Mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* 2018;50:1505–13.
33 Rosseel Y. lavaan: an R package for structural equation modeling. *J Stat Softw* 2011;48:1–36.
34 Bonder MJ, Kurilshikov A, Tigchelaar EF, *et al*. The effect of host genetics on the gut microbiome. *Nat Genet* 2016;48:1407–12.
35 Goodrich JK, Davenport ER, Beaumont M, *et al*. Genetic determinants of the gut microbiome in UK twins. *Cell Host Microbe* 2016;19:731–43.
36 Enattah NS, Sahi T, Savilahti E, *et al*. Identification of a variant associated with adult-type hypolactasia. *Nat Genet* 2002;30:233–7.
37 Wikoff WR, Anfora AT, Liu J, *et al*. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc Natl Acad Sci U S A* 2009;106:3698–703.
38 Dodd D, Spitzer MH, Van Treuren W, *et al*. A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites. *Nature* 2017;551:648–52.
39 Makki K, Deehan EC, Walter J, *et al*. The impact of dietary fiber on gut microbiota in host health and disease. *Cell Host Microbe* 2018;23:705–15.
40 Chaudhary P, Kumar NN, Deobagkar DN. The glucanases of Cellulomonas. *Biotechnol Adv* 1997;15:315–31.
41 Faith JJ, Guruge JL, Charbonneau M, *et al*. The long-term stability of the human gut microbiota. *Science* 2013;341:1237439.
42 Russell WR, Duncan SH, Scobbie L, *et al*. Major phenylpropanoid-derived metabolites in the human gut can arise from microbial fermentation of protein. *Mol Nutr Food Res* 2013;57:523–35.
43 Aragozzini F, Ferrari A, Pacini N, *et al*. Indole-3-lactic acid as a tryptophan metabolite produced by Bifidobacterium spp. *Appl Environ Microbiol* 1979;38:544–6.
44 Sedighi M, Razavi S, Navab-Moghadam F, *et al*. Comparison of gut microbiota in adult patients with type 2 diabetes and healthy individuals. *Microb Pathog* 2017;111:362–9.
45 Candela M, Biagi E, Soverini M, *et al*. Modulation of gut microbiota dysbioses in type 2 diabetic patients by macrobiotic Ma-Pi 2 diet. *Br J Nutr* 2016;116:80–93.
46 Wu X, Ma C, Han L, *et al*. Molecular characterisation of the faecal microbiota in patients with type II diabetes. *Curr Microbiol* 2010;61:69–78.
47 Wu H, Esteve E, Tremaroli V, *et al*. Metformin alters the gut microbiome of individuals with treatment-naive type 2 diabetes, contributing to the therapeutic effects of the drug. *Nat Med* 2017;23:850–8.
48 Gurung M, Li Z, You H, *et al*. Role of gut microbiota in type 2 diabetes pathophysiology. *EBioMedicine* 2020;51:102590.
49 Pedersen HK, Gudmundsdottir V, Nielsen HB, *et al*. Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* 2016;535:376–81.
50 Zhou W, Sailani MR, Contrepois K, *et al*. Longitudinal multi-omics of host-microbe dynamics in prediabetes. *Nature* 2019;569:663–71.