A Fecal Metabolite Signature of Impaired Fasting Glucose: Results From Two Independent Population-Based Cohorts

Ana Nogal,¹ Francesca Tettamanzi,^{1,2} Qiuling Dong,³ Panayiotis Louca,¹ Alessia Visconti,¹ Colette Christiansen,^{1,4} Taylor Breuninger,⁵ Jakob Linseisen,^{5,6,7} Harald Grallert,^{3,8} Nina Wawro,^{5,9} Francesco Asnicar,¹⁰ Kari Wong,¹¹ Andrei-Florin Baleanu,¹ Gregory A. Michelotti,¹¹ Nicola Segata,¹⁰ Mario Falchi,¹ Annette Peters,^{8,9,12} Paul W. Franks,^{13,14} Vincenzo Bagnardi,¹⁵ Tim D. Spector,¹ Jordana T. Bell,¹ Christian Gieger,^{3,8} Ana M. Valdes,¹⁶ and Cristina Menni¹

Prediabetes is a metabolic condition associated with gut microbiome composition, although mechanisms remain elusive. We searched for fecal metabolites, a readout of gut microbiome function, associated with impaired fasting glucose (IFG) in 142 individuals with IFG and 1,105 healthy individuals from the UK Adult Twin Registry (TwinsUK). We used the Cooperative Health Research in the Region of Augsburg (KORA) cohort (318 IFG individuals, 689 healthy individuals) to replicate our findings. We linearly combined eight IFGpositively associated metabolites (1-methylxantine, nicotinate, glucuronate, uridine, cholesterol, serine, caffeine, and protoporphyrin IX) into an IFG-metabolite score, which was significantly associated with higher odds ratios (ORs) for IFG (TwinsUK: OR 3.9 [95% CI 3.02-5.02], P < 0.0001, KORA: OR 1.3 [95% CI 1.16-1.52], P < 0.0001) and incident type 2 diabetes (T2D: TwinsUK: hazard ratio 4 [95% CI 1.97-8], P = 0.0002). Although these are host-produced metabolites, we found that the gut microbiome is strongly associated with their fecal levels (area under the curve >70%). Abundances of Faecalibacillus intestinalis, Dorea formicigenerans, Ruminococcus torques, and Dorea sp. AF24-7LB were positively associated with IFG, and such associations were partially mediated by 1-methylxanthine and nicotinate (variance accounted for mean 14.4% [SD 5.1], P < 0.05). Our results suggest that the gut microbiome is linked to prediabetes not only via the production of microbial metabolites but also by affecting intestinal absorption/excretion of host-produced metabolites and

¹Department of Twin Research, King's College London, St Thomas' Hospital Campus, London, U.K.

²Humanitas Clinical and Research Centre, IRCCS, Rozzano (Milan), Italy

- ³Institute of Epidemiology, Helmholtz Zentrum München, Research Unit of Molecular Epidemiology, German Research Center for Environmental Health (GmbH), Neuherberg, Germany
- ⁴School of Mathematics and Statistics, The Open University, Milton Keynes, U.K.
 ⁵Epidemiology, University Hospital Augsburg, University of Augsburg, Augsburg, Germany

ARTICLE HIGHLIGHTS

- Prediabetes is a metabolic condition associated with gut microbiome composition, although mechanisms remain elusive.
- We investigated whether there is a fecal metabolite signature of impaired fasting glucose (IFG) and the possible underlying mechanisms of action.
- We identified a fecal metabolite signature of IFG associated with prevalent IFG in two independent cohorts and incident type 2 diabetes in a subanalysis. Although the signature consists of metabolites of nonmicrobial origin, it is strongly correlated with gut microbiome composition.
- Fecal metabolites enable modeling of another mechanism of gut microbiome effect on prediabetes by affecting intestinal absorption or excretion of host compounds and xenobiotics.

xenobiotics, which are correlated with the risk of IFG. Fecal metabolites enable modeling of another mechanism of gut microbiome effect on prediabetes and T2D onset.

Type 2 diabetes (T2D) is a leading cause of mortality and morbidity (1), affecting >536.6 million people (10.5% of the total population) worldwide (2), thus representing a huge public health burden (1). The causation of T2D is multifactorial, influenced by host genetics and environmental

⁶ZIEL-Institute for Food & Health, Technische Universität München, Freising, Germany

⁷Institute for Medical Information Processing, Biometry, and Epidemiology, Medical Faculty, Ludwig-Maximilian University Munich, Munich, Germany

⁸German Center for Diabetes Research (DZD), Neuherberg, Germany

⁹Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH), Neuherberg, Germany

¹⁰Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Trento, Italy

factors, including diet, obesity, inactivity, and smoking, and the interaction between these factors (3). Furthermore, its onset is gradual, with people progressing through a state of prediabetes (4), and is defined as impaired levels of fasting glucose (IFG), and/or glucose intolerance, and/or elevated hemoglobin A_{1c} (Hb A_{1c}) (5).

Over the past decade, T2D and prediabetes have been linked by us and others (6-8) to changes in the gut microbiota, and we have recently demonstrated that T2D development is preceded by an alteration in gut microbiota composition (7). A critical challenge in human microbiome research, however, is to characterize and quantify metabolic activity across the full microbial ecosystem (9). The gut microbiome is highly variable, and different bacterial types may have similar metabolic effects on the host. Microbial metabolites are now widely seen as key mediators of the effects of gut microbiome composition on human physiology (10). Fecal metabolites provide a functional readout of the gut microbiome (11,12) and are a novel tool to explore links between gut microbiome composition and activity, host phenotypes, and heritable complex traits, thus improving our understanding of the impact that the gut microbiome can have on its host (11). As the gut microbiome is modifiable with nutritional and lifestyle interventions (13), it is of utmost importance to identify alterations in the fecal metabolites abundances, which reflect metabolic activity perturbations of the human gut microbial ecosystem that might lead to T2D onset.

In the first fecal metabolomics study of prediabetes to date, we aim to identify a fecal metabolite signature of this condition in two independent cohorts to shed light on mechanisms of action underlying T2D onset and development. Addressing this challenge also has long-term implications for future studies into therapies and lifestyle interventions that alter microbial metabolic activity to improve human health.

RESEARCH DESIGN AND METHODS

A flowchart of the study design with the main results is presented in Fig. 1.

Discovery Cohort

We analyzed data from 1,247 nonrelated individuals from UK Adult Twin Registry (TwinsUK) (14), for whom concurrent

- ¹²Munich Heart Alliance, German Center for Cardiovascular Research (DZHK e.V., Partner-Site Munich), Munich, Germany
- ¹³Lund University Diabetes Center, Lund University, Malmö, Sweden
- ¹⁴Department of Clinical Sciences, Lund University, Malmö, Sweden
- ¹⁵Department of Statistics and Quantitative Methods, University of Milan-Bicocca, Milan, Italy
- ¹⁶Academic Rheumatology Clinical Sciences Building, Nottingham City Hospital, University of Nottingham, U.K.

Corresponding authors: Cristina Menni, cristina.menni@kcl.ac.uk, and Ana M. Valdes, ana.valdes@nottingham.ac.uk

nontargeted fecal metabolomic profiling (526 metabolites at fasting) and glucose/diabetes information were available (cross-sectional design). Concurrent metagenome sequencing (as a measure of the gut microbiome composition) was also available for a subset of 342 individuals. Subjects were classified into three groups following the American Diabetes Association criteria based on isolated fasting glucose levels (15) at the time of the initial sampling and at subsequent visits (on average, 3.5 [SD 2.0] visits, 4.6 [SD 2.7] years apart): individuals with T2D (fasting glucose \geq 7 mmol/L or physician's letter confirming diabetes diagnosis), individuals with IFG (fasting glucose >5.5 to <7 mmol/L, not on diabetes medication), and subjects without IFG and T2D (fasting glucose >3.9 to \leq 5.5 mmol/L) (see Table 1). We refer to "healthy individuals" to indicate individuals without IFG and/or T2D.

Only one twin per twin pair was included in the analyses to eliminate potential bias through correlated error, which might inflate effect estimates.

In a small subanalysis, we included individuals with incident T2D (average follow-up time 2.1 [SD 1.3] years) and an independent subset of healthy individuals who remained healthy during follow-up.

All twins provided informed written consent and the study was approved by St Thomas' Hospital Research Ethics Committee (REC Ref: EC04/015).

Replication Cohort

The Cooperative Health Research in the Region of Augsburg (KORA) study is a population-based cohort study. The KORA FF4 study (2013–2014) is the second follow-up of KORA S4 (1999–2001). The 1,007 samples included in the study were collected in the morning between 8:00 A.M. and 10:30 A.M. after at least 8 h of fasting. Metabolon untargeted liquid chromatography/mass spectrometry (MS)-based techniques were applied to measure the metabolites in the KORA cohort (a different version of the platform used in TwinsUK). Healthy individuals and IFG individuals were assigned based on the same criteria as in TwinsUK (described in the above section and in Table 1).

Fecal Metabolomics Profiling

Metabolomics profiling was conducted using ultrahighperformance liquid chromatography-tandem MS (MS/MS) by the metabolomics provider Metabolon Inc. (Morrisville,

This article contains supplementary material online at https://doi.org/10.2337/ figshare.24114387.

A.N. and F.T. contributed equally.

¹¹Metabolon, Morrisville, NC



Figure 1 – Flowchart of the study design with the main results. Data, aims, methods, and results are shown in gray, blue, green, and pink squares, respectively. Mediation analyses were also performed for the metabolites making up the score that was predicted by the gut microbiome composition with an AUC >70%. Cov, covariates (age, BMI, and sex).

NC) on fecal samples from participants in the TwinsUK and KORA cohorts (Supplementary Material). The metabolomic data set measured by Metabolon includes 526 known metabolites for TwinsUK belonging to the following broad categories—amino acids, peptides, carbohydrates, energy intermediates, lipids, nucleotides, cofactors and vitamins, and xenobiotics—of which 357 were also measured in KORA. These include metabolites of established microbial origin (16). A complete list of the included metabolites with their superpathways, subpathways, Kyoto Encyclopedia of Genes and Genomes and Human Metabolome Database identifiers are reported in Supplementary Table 1. We imputed to the day minimum metabolites with <20% missing.

Metagenomic Assessment

Gut microbiota composition was generated from fecal shotgun metagenomes for a subset of the discovery cohort. DNA extraction, library preparation, and sequencing were conducted as detailed in Visconti et al. (11). For details see the Supplementary Material. Of note, gut microbiota composition is described by species-level genome bins (SGBs), which is the best proxy to define microbial species (17).

Statistical Analysis

Statistical analyses were conducted using R 4.2.2 software. To identify a fecal metabolite signature of prediabetes, we ran logistic regressions adjusting for age, BMI, sex, and multiple testing using the Benjamini and Hochberg method (18) (false discovery rate [FDR] <0.05). We then checked whether the metabolites significantly associated with IFG in the discovery set were also replicated in KORA (P < 0.1). We used a less stringent threshold for KORA because of the winner's curse (the effect sizes of the most strongly associated variables within a cohort-specific analysis are inflated) (19). Results were meta-analyzed using inverse-variance random-effect meta-analysis. We then created the IFG-metabolite score

			Discovery cohort	: TwinsUK			Rep	lication cohort:	KORA
	Prev	valent IFG ($n = 1$,247)	Inci	dent T2D (<i>n</i> =	27)	Pre	valent IFG (n = .	1,007)
	Healthy individuals	IFG individuals	Differences between groups (P value)	Healthy individuals	T2D individuals	Differences between groups (P value)	Healthy individuals	IFG individuals	Differences between groups (<i>P</i> value)
ADA definition (15), fasting glucose, mmol/L	≤5.5	>5.5 and <7	I	≤5.5	∠≂	I	≤5.5	>5.5 and $<$ 7	1
No.	1,105	142	I	17	10	I	689	318	I
Females, %	88.8	62	0.003	94.1	06	÷	58.1	35.5	
Age, years	56.6 (14.9)	67.1 (10)	<0.0001	66.5 (6.6)	65 (7.7)	0.7	55.2 (10.9)	59.8 (10.8)	<0.0001
BMI, kg/m ²	25.2 (4.6)	28.5 (5.1)	<0.0001	25.3 (3.3)	35.1 (6.7)	0.0004	26.1 (4.1)	28.4 (4.5)	<0.0001
Circulating fasting glucose, mmol/L	4.5 (0.3)	5.9 (0.4)	<0.0001	3.8 (0.3)	4.6 (1.5)	0.06	5.1 (0.3)	5.9 (0.3)	<0.0001
SBP, mmHg	125 (13.6)	134 (17)	<0.0001	132 (21.5)	133 (9.5)	0.8	114.5 (15.9)	123.1 (15.3)	<0.0001
DBP, mmHg	74.7 (8.1)	77.9 (10.3)	<0.0001	73.8 (11.9)	81.7 (8.1)	0.05	72.3 (8.9)	76 (9.7)	<0.0001
Circulating HDL, mmol/L	1.8 (1.2)	1.6 (1)	0.003	1.7 (0.4)	1.3 (0.2)	0.004	1.8 (0.5)	1.6 (0.5)	<0.0001
Circulating total cholesterol, mmol/L	4.1 (0.5)	4.1 (0.7)	0.73	4.7 (1.2)	3.6 (0.8)	0.02	5.6 (1)	5.7 (1)	0.008
Circulating triglycerides, mmol/L	1 (1)	1.6 (2.7)	0.0003	1 (0.3)	1.3 (0.4)	0.01	1.2 (0.7)	1.4 (0.9)	<0.0001
aHEI	70.5 (6.4)	70.1 (6.5)	0.49	72.8 (9.9)	71.4 (6.2)	0.68	NA	NA	NA
Current smoker, <i>n</i>	No: 1,060 Yes: 45	No: 139 Yes: 3	0.36	No: 17	No: 10	I	No: 346 Yes: 343	No: 139 Yes: 179	0.0
Activity levels, <i>n</i>	Low: 100 Moderate: 802 High: 203	Low: 13 Moderate: 102 High: 27	0.98	Low: 3 Moderate: 11 High: 3	Low: 2 Moderate: 5 High: 3	0.71	Inactive: 225 Active: 464	Inactive: 133 Active: 185	0.006
Continuous variables are presented healthy subjects from the IFG and in differences existed between the dit systolic blood pressure.	d as mean (SD).	Aeasures are sho sets. The <i>P</i> valu oups for the des	own at baseline. "He es are from a Wilco scribed parameters.	ealthy individual xon test/t test (c ADA, Americar	s" refers to in continuous var Diabetes As	dividuals with no iable) or χ^2 test (c sociation; DBP, d	IFG or T2D. Th ategorical varia iastolic blood p	iere was no ove ble), calculated vressure; NA, no	arlap between the to check whether available; SBP,

Table 1—Descriptive characteristics of the study populations

by linearly combining the replicated metabolites along with covariates. To assess the performance of the score in predicting prevalent IFG and incident T2D, we calculated the area under the curve (AUC) values obtained using fivefold cross-validation (caret package implemented in R [20]). Finally, logistic and Cox regressions were used to investigate the association between the IFG score (*Z*-scaled) and prevalent IFG risk and incident T2D risk, respectively.

Given the strong association between fecal metabolites and gut microbiome composition (12), we investigated to what extent the gut microbiota composition was associated with each of the replicated metabolites using random forest regressors and classifiers with compositional data and fivefold cross-validation. The performance was calculated using the average of the obtained Spearman correlations between the observed metabolite levels and the levels predicted by the model (denoted as ρ) over the fivefolds used as a test set for the regressors and the average of the obtained AUC values over the testing folds for the classifiers. For details see the Supplementary Material.

We further investigated the associations between their top 100 bacterial features and IFG by running logistic regression models adjusting for covariates and multiple testing species (FDR <0.05). Specifically, we included all of the fecal metabolites that could be predicted by the gut microbiome with an AUC >70%, and we then focused on those that had an outstanding prediction performance (AUC >90%).

Finally, we used formal mediation analysis as implemented in the R package "mediation" with 1,000 nonparametric bootstrap samples (21) to test the mediation effects of the metabolites on the total effect of the gut bacteria on IFG. The mediation model was used to quantify both the direct effect of these gut bacterial species on IFG and the indirect (mediated) effects mentioned above while controlling for age, BMI, and sex. The variance accounted for (VAF) score, which represents the ratio of indirect-to-total effect and determines the proportion of the variance explained by the mediation process, was used to determine the significance of the mediation effect.

Data and Resource Availability

The data used in this study are held by the Department of Twin Research at King's College London. The data can be released to bona fide researchers using our normal procedures overseen by the Wellcome Trust and its guidelines as part of our core funding (https://twinsuk.ac.uk/resourcesfor-researchers/access-our-data/). The gut microbiome data are available on EBI (https://www.ebi.ac.uk/) under accession number PRJEB32731 (TwinsUK). The KORA FF4 datasets are available upon application through the KORA-PASST (project application self-service tool, https://www.helmholtzmunich.de/epi/research/cohorts/kora-cohort/data-use-andaccess-via-korapasst/index.html).

RESULTS

We included 1,247 unrelated individuals from the TwinsUK cohort who had fecal metabolite measures along with

glucose/diabetes and prediabetes information. Of these, 142 individuals had IFG (mean fasting glucose 5.9 mmol/L [SD 0.4]) and 1,105 were healthy individuals (mean fasting glucose 4.5 mmol/L [SD 0.3]). Descriptive characteristics of the discovery and replication populations are included in Table 1.

Fecal Metabolites Associated Cross-sectionally to IFG

Of the 526 known fecal metabolites analyzed in TwinsUK, the fecal abundances of 26 compounds were associated with IFG after adjusting for age, BMI, sex, and multiple testing (FDR <0.05) (Fig. 2). Identified metabolites were mainly amino acids (n = 7) and lipids (n = 7), but also included xenobiotics (n = 4), cofactors and vitamins (n = 3), nucleotides (n = 2), carbohydrates (n = 2), and one energyrelated metabolite (Fig. 2). All significant metabolites, but 3-hydroxyoleate, octadecanedioate (C18-DC), azelate (C9-DC), y-tocotrienol, and enterolactone, were positively associated with IFG (Fig. 2). Of the 26 metabolites, 18 were also measured in KORA (Supplementary Table 1), and 8 metabolites were replicated (P < 0.1) (Fig. 3). These were the lipid cholesterol (sterol metabolism), the carbohydrate glucuronate (aminosugar metabolism), the cofactors/vitamins nicotinate (nicotinate and nicotinamide metabolism) and protoporphyrin IX (hemoglobin and porphyrin metabolism), the xenobiotics caffeine and 1-methylxanthine (both involved in the xanthine metabolism), the amino acid serine (glycine, serine, and threonine metabolism), and the nucleotide uridine (pyrimidine metabolism). The correlation matrices for the eight fecal metabolites in TwinsUK and KORA are depicted in Supplementary Fig. 1. We combined the results from both cohorts using inverse-variance randomeffect meta-analysis (Fig. 3).

IFG-Metabolite Score and Predictive Power

We then generated the IFG-metabolite score using TwinsUK individuals:

 $IFG-metabolite \ score = -8.79 + 0.07 \times glucuronate + 0.25 \times protoporphyrin \ IX + 0.09 \times 1-methylxanthine + 0.14 \times cholesterol + 0.04 \times serine + 0.07 \times uridine + 0.04 \times nicotinate + 0.17 \times caffeine + 0.07 \times age + 0.1 \times BMI - 0.6 \times sex (female = 1)$

The IFG-metabolite score was associated with an increased risk of IFG in TwinsUK (odds ratio [OR] 3.9 [95% CI 43.02–5.02], P < 0.0001) and in KORA (OR 1.3 [95% CI 1.16–1.52], P < 0.0001). The association remained significant when further adjusting for clinical covariates (i.e., systolic and diastolic blood pressure, circulating levels of HDL, total cholesterol, and triglycerides, alternative health eating index [aHEI – not available in KORA], activity levels and smoking status) (Table 1) in both cohorts (TwinsUK: OR 3.4 [95% CI 2.65–4.49], P < 0.0001; KORA: OR 1.2 [95% CI 1.06–1.41], P = 0.008). Finally, the IFG-metabolite score accurately predicted prevalent IFG in TwinsUK with an AUC of 79.8% (95% CI 76.3–83.3) in fivefold stratified cross-validation and



Figure 2—Fecal metabolites significantly associated with IFG in 1,247 individuals from TwinsUK after adjusting for baseline age and BMI, sex, and multiple testing (FDR <0.05). Bars represent the OR. Base labels illustrate subpathways. met., metabolism.

outperformed the model including only covariates (AUC 77.2% [95% CI = 73.6–81]) by 2.6% (Δ 95% CI 2.7–2.1). In KORA, the IFG-metabolite score (top vs. lowest decile) could satisfactory predict prevalent IFG (AUC 65.4 [95% CI 57.9–73]).

Subanalysis: Incident T2D

In a small independent sample from TwinsUK (descriptive characteristics are shown in Table 1) consisting of 17 healthy individuals (different from the healthy subjects of the IFG data set) and 10 individuals with incident T2D (follow-up time between fecal metabolite measurements and incident events: mean 2.1 [SD 1.3] years), the IFG-metabolite score was also predictive of an increased risk of incident T2D (hazard ratio 4 [95% CI 1.97–8], P = 0.0002) in TwinsUK after further adjusting for baseline circulating glucose levels. It also accurately predicted incident T2D (AUC 83.3% [95% CI 74.4–92.2]), while a model using baseline circulating glucose levels as predictor presented a lower prediction power (AUC 72.4% [95% CI 51.8–92.9]).

Gut Microbiome–Fecal Metabolites Association

We further evaluated the extent to which the gut microbiota was associated with the fecal abundances of the eight replicated metabolites using the AUC obtained by the random forest classifiers and the Spearman correlations (denoted as ρ) between the real abundances and predicted values by the random forest regressors. We included a subset of 342 individuals from TwinsUK with concurrent gut microbiota composition assessed by shotgun metagenomics and fecal metabolites measurements. Descriptive characteristics of this subset are shown in Supplementary Table 2.

The gut microbiome composition was strongly associated with the replicated metabolites, with performance metric values ranging from an AUC of 70.7% (95% CI, 69.1–72.4) and ρ = 0.24 (95% CI, 0.23–0.25) for caffeine to an AUC of 91.4% (95% CI, 90.8–91.9) and ρ = 0.62 (95% CI, 0.62–0.62) for 1-methylxanthine (Fig. 4A and Supplementary Table 3). Protoporphyrin IX was the only metabolite presenting a moderate association (AUC 64.8% [95% CI 63.9–65.6]; ρ = 0.25 [95% CI 0.24–0.26]) (Fig. 4A).

We then investigated whether the abundances from their top 100 bacterial features based on the random



Figure 3—Fecal metabolites significantly associated with IFG after adjusting for age, BMI, and sex in TwinsUK (FDR <0.05), KORA (P < 0.1) and in the overall cohort (applying inverse-variance random-effect meta-analysis). The OR and 95% CI are indicated.

forest models were also significantly associated with IFG (Supplementary Table 4). We focused on the fecal metabolites that presented the strongest associations with the gut microbiome composition (AUC >90%—outstanding prediction performance; 1-methylxathine and nicotinate). We identified four characterized gut bacterial species for 1-methylxanthine and nicotinate, of which three overlapping (overlapping: Dorea formicigenerans, Ruminococcus torques, and Faecalibacillus intestinalis; 1-methylxanthine only: Dorea sp. AF24-7LB; nicotinate only: Dorea sp. AF36-15AT), that were positively associated with IFG after adjusting for age, BMI and sex (FDR <0.05) (Supplementary Table 4). We, therefore, performed a formal mediation analysis adjusting for age, BMI, and sex to determine whether 1-methylxanthine and/or nicotinate mediated the associations between these species

and IFG. The analysis revealed that 1-methylxanthine acted as a potential mediator in the positive associations of Dorea sp. AF24-7LB (VAF = 10.3%, P = 0.03) and R. torques (VAF = 9.7%, P = 0.04) with IFG, while nicotinate acted as a potential mediator in the positive associations of F. intestinalis (VAF = 22.3%, P = 0.002), D. formicigenerans (VAF = 15.8%, P = 0.002), and R. torques (VAF = 14.1%, P = 0.03) with IFG (Fig. 4B). We further ran mediation analyses for the metabolites that could be predicted by the gut microbiome with an AUC >70%. As reported in Supplementary Fig. 2, uridine, serine, cholesterol, and caffeine were also mediators in the associations between different species (e.g., Dorea spp. and Anaerobutyricum hallii) and IFG. Models were not further adjusted for other comorbidities (e.g., systolic and diastolic blood pressure, circulating levels of HDL, total cholesterol and triglycerides, aHEI, activity



Figure 4—Associations of the gut microbiota with the eight fecal replicated metabolites and IFG in 342 TwinsUK participants. *A*: Influence of the gut microbiota composition in the fecal abundances of the eight replicated metabolites estimated by random forest regressors (Spearman correlations between the real value of each metabolite and the value predicted) and classifiers (AUC). Red and blue bars represent the mean AUC and Spearman correlations with the respective 95% CIs across fivefolds, respectively. *B*: Mediation analyses of the associations between characterized gut bacterial species and IFG. Models were adjusted for age, BMI, and sex. Path coefficients are shown beside each path, and indirect effects and VAF score are indicated below each mediator (left: nicotinate, right: 1-methylxanthine). Only metabolites with a predictive power of AUC >90% in *A* are shown.

levels, and smoking status) as these were not significantly associated with the identified bacterial species or with the metabolites making up the score (Supplementary Table 5).

DISCUSSION

Here we identify for the first time a fecal metabolite signature of IFG that is associated with prevalent IFG in two independent cohorts and is also predictive of incident T2D in a small subanalysis. The fecal metabolites making up the score are not microbial-derived metabolites but are "host metabolites" (e.g., xenobiotics, cofactors, and vitamins). However, the gut microbiome can accurately predict their fecal abundances (AUC >70%). It is well known that the gut microbiome composition can affect diseases via several mechanisms (22). Circulating microbial metabolites have been reported by us and others to be reflective of gut microbiome diversity and composition (6-8) and predictive of prevalent and incident T2D (7). Taken together, this suggests that the gut microbiome can influence T2D, not only by producing metabolites that enter the bloodstream (7) but also by regulating the absorption or excretion of host-produced compounds, thereby influencing IFG and T2D risk. This hypothesis is further supported by the results of our mediation analysis showing that metabolites making up the score act as partial mediators on the significant associations between several gut microbial species, (e.g., F. intestinalis, D. formicigenerans, R. torques, and Dorea sp. AF24-7LB) and IFG.

Studies have shown that gut microbiome composition differs between individuals with prediabetes/diabetes and healthy subjects (6,7), with compositional shifts correlated with synthesis profile changes of gut bacteria-derived metabolites, including short-chain fatty acids, indolepropionic acid, and trimethylamine (7,22). These "microbial" metabolites enter into the bloodstream and reach different tissues, where they can influence glucose homeostasis and insulin resistance by activating or inhibiting signaling pathways (22). Nevertheless, the identified signature of prediabetes in this study consists of eight metabolites of nonbacterial origin. Serine is a nonessential amino acid mainly obtained by intrinsic synthesis (23). Glucuronate is a sugar acid derived from glucose and involved in the detoxification of xenobiotic compounds (24). Protoporphyrin IX is a cofactor ubiquitously present in the human body as a heme precursor (25). Nicotinate, also known as vitamin B3 and niacin, is a watersoluble vitamin that can be produced by the human body from tryptophan (26). Cholesterol, which is mainly produced by the liver, is an essential lipid of eukaryotic cell membranes and is also a precursor of bile acids and steroid hormones (27). Uridine is a necessary pyrimidine nucleotide for RNA synthesis produced by several reversible reactions (e.g., dephosphorylation of uridine monophosphate, deamination of a cytidine or combination of uracil and ribose 1-phosphate) (28). Caffeine and 1-methylxanthine are xenobiotics involved in the caffeine metabolism pathway (29).

Strikingly, we find that the gut microbiome is strongly associated with fecal levels of these metabolites, suggesting that the gut microbiome influences the absorption or excretion of compounds involved in various metabolic pathways (e.g., cholesterol, uridine, and glucuronate) and xenobiotics (e.g., caffeine and its derivatives), among others, and such levels of absorption or excretion are directly related to IFG. Our findings lead us to speculate that individuals with prediabetes present gut microbiome composition perturbations, which likewise influence the absorption or excretion of the identified compounds. This is further supported by the mediation analyses, which suggest that the associations between specific gut microbial species, including F. intestinalis, D. formicigenerans, R. torques, and Dorea sp. AF24-7LB, and IFG are mainly reflecting the effect of the gut microbiome in the absorption or excretion of the found compounds.

Under normal conditions, the small intestine can break down, emulsify, and absorb most nutrients, including fats, simple carbohydrates, and proteins (30). For instance, <5 g/day of fat are not absorbed and reach the colon (30). Nonetheless, the absorption capability of the gut can be limited depending on the gut microbiome composition (31). A study conducted by Basolo et al. (31) demonstrated that changes in participants' gut microbiome composition, due to diet or antibiotic use, impaired nutrient absorption. Several mechanisms might explain how gut microbiome composition might influence absorption, and thus, the disease onset (32-34). For instance, the gut microbiome can affect the gut barrier, which consists of a collection of physical and chemical structures that protect the host from pathogenic invasions and harmful stimuli (32). This can be provoked by the presence of pathogen-associated molecular patterns, such as lipopolysaccharides, in the cell walls of some gram-negative bacteria, which play an important role in intestinal absorption, blood glucose, and inflammation (33). Moreover, changes in the permeability of the gut barrier can be caused by an unbalanced increase in bacteria able to degrade mucin (the main component of mucus, which covers the epithelial surfaces of the gastrointestinal tract) (32). Indeed, in this study, we identify that individuals with prediabetes present larger abundances compared with healthy individuals of the mucindegraders D. formicigenerans (35) and R. torques (36), which have been previously associated with lower nutrient absorption (36). Finally, some gut microbes can also reduce absorption in the jejunum by altering the expression of intestinal transporters of different types of compounds (34).

Another possible explanation for our findings could be a reduction of specific beneficial bacteria able to use these compounds, thus resulting in increased excretion (27,37). In the case of cholesterol, bacterial members of the genera *Bifidobacterium, Lactobacillus,* and *Peptostreptococcus* are needed to convert cholesterol into coprostanol (27). Likewise, an inefficient cholesterol-coprostanol conversion is linked to cardiometabolic diseases (27). For glucuronate, most of it is not absorbed by the small intestine; however,

under normal conditions, the amounts that make it to the colon are then efficiently used by *Bifidobacterium* (37).

This work has several strengths. Our study benefits from a large, accurately phenotyped discovery cohort with metabolomic profiling and gut microbiome composition. We were also able to replicate our findings in a large independent cohort, thus strengthening our findings. Finally, a machine learning algorithm was applied to investigate the prediction of the gut microbiota to the levels of the found eight metabolites, allowing us to simultaneously integrate all the species in the models.

We also note some study limitations. First, the crosssectional nature of the data used for our primary analysis does not allow us to determine the temporal link between IFG and the identified fecal metabolites.

Second, HbA_{1c} , postprandial glucose to derive impaired glucose tolerance, which more closely resembles the T2D state (38), and a clinician's diagnosis were not available in the discovery cohort. Thus, the division of categories in this study is derived from IFG.

Third, the sample size for the subanalysis looking at incident T2D was limited, and we were unable to seek independent replication as, to the best of our knowledge, there are no other cohorts in the world that have measured this fecal metabolome panel and incident T2D. Future studies with larger sample size are therefore needed to test the robustness of the IFG metabolite score to predict incident T2D.

Fourth, there was not a full overlap between metabolites measured in the discovery and validation data sets, which might cause the loss of metabolites of interest to study.

Fifth, the included study groups were unbalanced in age and sex. Hence, although we adjusted all analyses for them and other important clinical variants, the confidence of the results is lowered. In addition, gut microbiota composition data were only available for a subset from the discovery set, and therefore, we could not replicate the mediation analysis in KORA. Furthermore, the Spearman correlations between the predicted (from gut microbiome composition) and actual levels of the metabolites were modest. Indeed, random forest models were trained based on microbial features extracted from metagenomic data, which does not retrieve all species present in a microbiome sample for procedural and technical reasons.

Finally, this study does not include measures of permeability markers, which would contribute to a better understanding of the role of intestinal permeability in the absorption or excretion of the identified compounds.

In conclusion, we are proposing a novel mechanism of how gut microbiome composition affects prediabetes and, consequently, the onset of T2D. The gut microbiome is linked to prediabetes not only by microbial-derived metabolites but also by affecting intestinal absorption or excretion of metabolites of nonmicrobial origin, which are correlated with the risk of IFG and incident T2D. Henceforth, to better understand the onset of T2D, the effect of the gut microbiome in the excretion and/or absorption of host-produced compounds and xenobiotics also needs to be also considered.

Acknowledgments. The authors thank all the participants of TwinsUK for contributing and supporting their research and all participants of KORA for their long-term commitment to the KORA study, the staff for data collection and research data management, and the members of the KORA Study Group (https://www.helmholtz-munich.de/en/epi/cohort/kora) who are responsible for the design and conduct of the study. Finally, the authors thank Caroline Le Roy (Department of Twin Research & Genetic Epidemiology, King's College, London) for the helpful discussion. Data collection in the KORA study is done in cooperation with the University Hospital of Augsburg.

Funding. This research was funded by the Chronic Disease Research Foundation, and in part by the Wellcome Trust (grant no. 212904/Z/18/Z) and by DiabetesUK (19/0006053). TwinsUK receives funding from the Wellcome Trust, the European Commission H2020 grants SYSCID (contract #733100), the National Institute for Health Research (NIHR) Clinical Research Facility, and the Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London, the Chronic Disease Research Foundation, the UKRI Medical Research Council (MRC)/British Heart Foundation Ancestry and Biological Informative Markers for Stratification of Hypertension (AIM-HY, MR/M016560/1). This work was also supported by UKRI grant MR/ W026813/1 to C.M. and A.M.V. The KORA study was initiated and financed by the Helmholtz Zentrum München-German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Stool sample collection and metabolomics analysis in KORA FF4 was supported by iMED, a research alliance within the Helmholtz Association, Germany. C.M., L.P., and A.N. are funded by the Chronic Disease Research Foundation. A.M.V. is supported by the National Institute for Health Research Nottingham Biomedical Research Centre.

Duality of Interest. This study received support from Zoe Ltd. T.D.S. is a cofounder and shareholder of Zoe Ltd. A.M.V., P.W.F., F.A., and N.S. are consultants to Zoe Ltd. K.W. and G.A.M. are employees of Metabolon Inc. No other potential conflicts of interest relevant to this article were reported.

Author Contributions. A.N., F.T., Q.D., A.V., C.C., H.G., F.A., and C.G. curated the data. A.N., F.T., Q.D., and C.M. performed the formal analyses. A.N., A.M.V., and C.M. wrote the manuscript. Q.D., P.L., A.V., C.C., T.B., J.L., H.G., N.W., F.A., K.W., A.-F.B., G.A.M., N.S., M.F., A.P., P.W.F., V.B., T.D.S., J.T.B., and C.G. contributed reagents/materials/analysis tools. T.D.S. and C.M. conceived and designed the experiments. All of the authors revised the manuscript. C.M. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol 2018;14:88–98

 Sun H, Saeedi P, Karuranga S, et al. IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022;183:109119

3. Kolb H, Martin S. Environmental/lifestyle factors in the pathogenesis and prevention of type 2 diabetes. BMC Med 2017;15:131

 Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 2002;346:393–403

5. Elliott TL, Pfotenhauer KM. Classification and diagnosis of diabetes. Prim Care 2022;49:191–200

6. Aydin Ö, Nieuwdorp M, Gerdes V. The gut microbiome as a target for the treatment of type 2 diabetes. Curr Diab Rep 2018;18:55

7. Menni C, Zhu J, Le Roy Cl, et al. Serum metabolites reflecting gut microbiome alpha diversity predict type 2 diabetes. Gut Microbes 2020;11: 1632–1642

8. Zhang Z, Tian T, Chen Z, Liu L, Luo T, Dai J. Characteristics of the gut microbiome in patients with prediabetes and type 2 diabetes. PeerJ 2021; 9:e10952

9. Maurice CF, Turnbaugh PJ. Quantifying the metabolic activities of humanassociated microbial communities across multiple ecological scales. FEMS Microbiol Rev 2013;37:830–848

10. Krautkramer KA, Fan J, Bäckhed F. Gut microbial metabolites as multikingdom intermediates. Nat Rev Microbiol 2021;19:77–94

11. Visconti A, Le Roy Cl, Rosa F, et al. Interplay between the human gut microbiome and host metabolism. Nat Commun 2019;10:4505

12. Zierer J, Jackson MA, Kastenmüller G, et al. The fecal metabolome as a functional readout of the gut microbiome. Nat Genet 2018;50:790-795

13. Conlon MA, Bird AR. The impact of diet and lifestyle on gut microbiota and human health. Nutrients 2014;7:17-44

14. Verdi S, Abbasian G, Bowyer RCE, et al. TwinsUK: the UK adult twin registry update. Twin Res Hum Genet 2019;22:523–529

 ElSayed NA, Aleppo G, Aroda VR, et al.; American Diabetes Association.
 Classification and diagnosis of diabetes: *Standards of Care in Diabetes*— 2023. Diabetes Care 2023;46(Suppl. 1):S19–S40

16. Donia MS, Fischbach MA. Human microbiota. Small molecules from the human microbiota. Science 2015;349:1254766

17. Pasolli E, Asnicar F, Manara S, et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. Cell 2019;176:649–662.e620

18. Thissen D, Steinberg L, Kuang D. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. J Educ Behav Stat 2002;27:77–83

19. Tugwell P, Knottnerus JA. A statistic to avoid being misled by the "winners curse". J Clin Epidemiol 2018;103:vi-viii

20. Kuhn M. Building predictive models in R using the caret package. J Stat Softw 2008;28:1-26

21. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R package for causal mediation analysis. J Stat Softw 2014;59:1–38

22. Nogal A, Valdes AM, Menni C. The role of short-chain fatty acids in the interplay between gut microbiota and diet in cardio-metabolic health. Gut Microbes 2021;13:1–24

23. Jiang J, Li B, He W, Huang C. Dietary serine supplementation: friend or foe? Curr Opin Pharmacol 2021;61:12–20

24. Fujiwara R, Yoda E, Tukey RH. Species differences in drug glucuronidation: Humanized UDP-glucuronosyltransferase 1 mice and their application for predicting drug glucuronidation and drug-induced toxicity in humans. Drug Metab Pharmacokinet 2018;33:9–16

25. Sachar M, Anderson KE, Ma X. Protoporphyrin IX: the good, the bad, and the ugly. J Pharmacol Exp Ther 2016;356:267–275

26. Moffett JR, Namboodiri MA. Tryptophan and the immune response. Immunol Cell Biol 2003;81:247–265

27. Kriaa A, Bourgin M, Potiron A, et al. Microbial impact on cholesterol and bile acid metabolism: current status and future prospects. J Lipid Res 2019;60: 323–332

28. Urasaki Y, Pizzorno G, Le TT. Uridine affects liver protein glycosylation, insulin signaling, and heme biosynthesis. PLoS One 2014;9:e99728

29. Keijzers GB, De Galan BE, Tack CJ, Smits P. Caffeine can decrease insulin sensitivity in humans. Diabetes Care 2002;25:364–369

30. de Vos WM, Tilg H, Van Hul M, Cani PD. Gut microbiome and health: mechanistic insights. Gut 2022;71:1020-1032

31. Basolo A, Hohenadel M, Ang QY, et al. Effects of underfeeding and oral vancomycin on gut microbiome and nutrient absorption in humans. Nat Med 2020;26:589–598

32. Raimondi S, Musmeci E, Candeliere F, Amaretti A, Rossi M. Identification of mucin degraders of the human gut microbiota. Sci Rep 2021;11:11094

 Anhê FF, Barra NG, Cavallari JF, Henriksbo BD, Schertzer JD. Metabolic endotoxemia is dictated by the type of lipopolysaccharide. Cell Rep 2021;36:109691
 Depommier C, Van Hul M, Everard A, Delzenne NM, De Vos WM, Cani PD. Pasteurized *Akkermansia muciniphila* increases whole-body energy expenditure and fecal energy excretion in diet-induced obese mice. Gut Microbes 2020;11:1231–1245

Vacca M, Celano G, Calabrese FM, Portincasa P, Gobbetti M, De Angelis
 M. The controversial role of human gut lachnospiraceae. Microorganisms 2020;
 8:573

36. Kaczmarczyk M, Löber U, Adamek K, et al. The gut microbiota is associated with the small intestinal paracellular permeability and the development of the immune system in healthy children during the first two years of life. J Transl Med 2021;19:177

37. Asano T, Yuasa K, Kunugita K, Teraji T, Mitsuoka T. Effects of gluconic acid on human faecal bacteria. Microb Ecol Health Dis 1994;7:247–256

 Unwin N, Shaw J, Zimmet P, Alberti KG. Impaired glucose tolerance and impaired fasting glycaemia: the current status on definition and intervention. Diabet Med 2002;19:708–723