Title

Nontargeted and Targeted Metabolomic Profiling Reveals Novel Metabolite Biomarkers of Incident Diabetes in African Americans

Journal

Diabetes, 71(11)

ISSN

0012-1797

Authors

Chen, Zsu-Zsu
Pacheco, Julian Avila
Gao, Yan
et al.

Publication Date

2022-11-01

DOI

10.2337/db22-0033

Peer reviewed

Nontargeted and Targeted Metabolomic Profiling Reveals Novel Metabolite Biomarkers of Incident Diabetes in African Americans

Zsu-Zsu Chen,[1,2] Julian Avila Pacheco,[3] Yan Gao,[4] Shuliang Deng,[5] Bennet Peterson,[5] Xu Shi,[5] Shuning Zheng,[5] Usman A. Tahir,[2,5] Daniel H. Katz,[5] Daniel E. Cruz,[2,5] Debby Ngo,[2,6] Mark D. Benson,[2,5] Jeremy M. Robbins,[2,5] Xiuqing Guo,[7] Magdalena del Rocio Sevilla Gonzalez,[2,3,8] Alisa Manning,[2,3,8] Adolfo Correa,[4] James B. Meigs,[2,3,8] Kent D. Taylor,[7] Stephen S. Rich,[9] Mark O. Goodarzi,[10] Jerome I. Rotter,[7] James G. Wilson,[5] Clary B. Clish,[3] and Robert E. Gerszten[2,3,5]

Diabetes 2022;71:2426–2437 | https://doi.org/10.2337/db22-0033

Nontargeted metabolomics methods have increased potential to identify new disease biomarkers, but assessments of the additive information provided in large human cohorts by these less biased techniques are limited. To diversify our knowledge of diabetes-associated metabolites, we leveraged a method that measures 305 targeted or "known" and 2,342 nontargeted or "unknown" compounds in fasting plasma samples from 2,750 participants (315 incident cases) in the Jackson Heart Study (JHS)—a community cohort of self-identified African Americans—who are underrepresented in omics studies. We found 307 unique compounds (82 known) associated with diabetes after adjusting for age and sex at a false discovery rate of <0.05 and 124 compounds (35 known, including 11 not previously associated) after further adjustments for BMI and fasting plasma glucose. Of these, 144 and 68 associations, respectively, replicated in a multiethnic cohort. Among these is an apparently novel isomer of the 1-deoxyceramide Cer(m18:1/24:0) with functional geonomics and high-resolution mass spectrometry. Overall, known and unknown metabolites provided complementary information (median correlation ρ = 0.29), and their inclusion with clinical risk factors improved diabetes prediction modeling. Our findings highlight the importance of including nontargeted metabolomics methods to provide new insights into diabetes development in ethnically diverse cohorts.

More than 10% of the U.S. adult population has diabetes (1), and an additional 34% are at risk. Unfortunately, specific racial and ethnic groups, including African Americans (AAs), are disproportionately affected (2). This contributes to the clinical challenge of correctly determining individual type 2 diabetes (T2D) risk (3,4), which is important for disease prevention. Dysglycemia, dyslipidemia (5), obesity (6), and genetic polymorphisms (7–10) are known risk factors, but questions remain about how they interplay to cause disease. High-throughput profiling of circulating small molecules—known as metabolomics—has identified T2D biomarkers in large human cohorts and nominated potential causal pathways for further study (11–20). A majority of the published data, however, have used targeted methods that focus on a group of mass spectrometry (MS) peaks that have been chemically annotated and are referred to as "known"

[1]Division of Endocrinology, Diabetes, and Metabolism, Beth Israel Deaconess Medical Center, Boston, MA
[2]Harvard School of Medicine, Boston, MA
[3]Broad Institute of MIT and Harvard, Boston, MA
[4]University of Mississippi Medical Center, Jacksonville, MS
[5]Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA
[6]Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA
[7]The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA
[8]Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA
[9]University of Virginia School of Medicine, Charlottesville, VA
[10]Division of Endocrinology, Diabetes, and Metabolism, Cedars-Sinai Medical Center, Los Angeles, CA

Corresponding author: Robert E. Gerszten, rgerszte@bidmc.harvard.edu

Received 19 January 2022 and accepted 26 July 2022

This article contains supplementary material online at https://doi.org/10.2337/figshare.20510979.

Z.-Z.C. and J.A.P. are co-first authors.

metabolites. This biases discovery toward compounds that participate in highly studied physiologic pathways and represent only a modest percentage of the circulating metabolome (21). Most cohorts studied are also of White individuals, frequently of European ancestry, despite the higher burden of disease in other race and ethnic groups (1) and the potential for differences in metabolite associations after race/ethnicity stratification (19).

In this study, we leveraged a hybrid liquid-chromatography MS (LC-MS) method to identify both targeted and nontargeted circulating compounds associated with diabetes in the Jackson Heart Study (JHS), a large community cohort of self-reported AA individuals (22). We replicated our findings in the Multi-Ethnic Study of Atherosclerosis (MESA) (23). Furthermore, we integrated these associations with available whole-genome sequencing (WGS) data, uncovering genetic variants in specific enzyme or solute carriers linked to these compounds that help inform unknown chemical identification. We first used this technique to identify dimethylguanidino valeric acid, a molecular marker of liver fat that was associated in genome-wide association studies with alanine glyoxylate aminotransferase 2 (AGXT2) (25). Here, we leveraged a similar technique coupled with innovative high-resolution and accurate MS to identify a novel metabolite marker of diabetes. Finally, we evaluated both the targeted and nontargeted compounds as clinical prediction biomarkers. These findings serve to diversify our understanding of circulating metabolites associated with diabetes and highlight potentially novel disease pathways.

RESEARCH DESIGN AND METHODS

Study Populations

The JHS is a community cohort of 5,306 self-identified AA individuals residing in Jackson, Mississippi, with detailed study design previously published (23). Diabetes status was assessed at examinations in 2000–2004, 2005–2008, and 2009–2013. Fasting plasma samples from 2,750 participants were profiled (1,159 individuals were selected from nested case-control studies for coronary heart disease and chronic kidney disease and 1,591 were randomly sampled from the remaining participants). Of these, 710 individuals had diabetes at baseline. An additional 315 developed diabetes after a mean follow-up of 10.2 years.

MESA is a U.S. community-based cohort study that recruited individuals who self-identified as White, AA, Hispanic, or Chinese American (24). At the baseline examination (2000–2002), 918 individuals were free of diabetes and underwent metabolomics profiling (403 self-identified as White, 175 as AA, 268 as Hispanic, and 72 as Chinese American). All individuals were included in the replication cohort to improve statistical power. During a mean follow-up of 8.9 years, 126 individuals developed diabetes.

Written consent was obtained from all of the study participants, and study protocols were approved by the institutional review boards of Beth Israel Deaconess Medical Center and each JHS and MESA study site.

Clinical Variables and Outcome

Diabetes was defined in JHS at each examination as a fasting plasma glucose (FPG) $\geq$126 mg/dL, hemoglobin A_{1c} (HbA_{1c}) $\geq$6.5%, diabetes diagnosis, or diabetes medication use. Hypertension was defined as a systolic blood pressure (SBP) >140 mmHg, diastolic blood pressure (DBP) >90 mmHg, or use of hypertension medications. HbA_{1c}, FPG, insulin, and lipids were measured using standard laboratory techniques (25). HOMA for insulin resistance (HOMA-IR) was calculated using fasting insulin × FPG/22.5. The Chronic Kidney Disease Epidemiology Collaboration equation was used to calculate estimated glomerular filtration rate (eGFR) (26). In MESA, diabetes was defined as an FPG $\geq$126 mg/dL and/or use of diabetes medications, including insulin (27).

Overview of Metabolite Profiling

Fasting plasma samples were obtained at the baseline examination. A total of 2,649 LC-MS peaks—including targeted and nontargeted features—were measured by using two different LC-MS methods (hydrophilic interaction liquid chromatography [HILIC] positive and amide negative) that have been previously described (28,29). Quality control (QC) pools created by combining small-volume aliquots from all JHS samples were inserted every 20 samples and used to normalize intensity trends across batches and to calculate the coefficient of variation (CV) for each metabolite. Normalization was visually confirmed with plotted pre- and postnormalized data. The median CV was 4.0% for the targeted HILIC-positive method, 11.4% for nontargeted, and 6.9% for the targeted amide-negative method. More than 97% of the measured LC-MS features had <20% missingness.

Tandem MS Methods for Nontargeted Metabolite Feature Identification

A comprehensive tandem MS (MS/MS) library of all measured features was created using HILIC chromatography coupled to a Thermo ID-X Mass Spectrometer (Thermo Fisher Scientific, Waltham, MA) scanning in positive ion mode with different collision energies (10, 25, and 50 V) of study QC pools. To improve detection of low abundant features, QC pools were concentrated 10-fold. MS/MS data extraction was then conducted by scanning for precursors within ± 0.2 atomic mass units of the targeted feature and ± 0.1 min from the apex of the MS/MS detected peak. Parsed MS/MS was formatted for molecular structure predictions (*.ms) and loaded into SIRIUS+CSI:-Finger ID version 4.7.2 (30) with molecular formula predictions based on Orbitrap-specific settings (MS/MS isotope scorer: ignore; mass deviation: 5 ppm; candidates: 10; candidates per ion: 1; possible ionizations: [M+H]+, [M+K]+, and [M+Na]+). All databases were searched, including adducts [M+H]+, [M+K]+, and [M+Na]+, and the top three predicted chemical structure/compound identifications were exported.

Metabolomics Data Processing and Statistical Analyses With Clinical Traits and Outcomes

Nontargeted LC-MS peaks can represent adducts, dehydration products, or daughter ions of a parent compound. The 2,342 measured peaks were statistically reduced using a correlation matrix that clustered potential adducts and daughter ions with parent ions based on feature retention time (RT) and mass-to-charge (*m/z*) values. Only those designated as "primary features"—which we believe represent truly unique compounds (n = 1,434)—were included in the nontargeted analyses.

Compound concentration correlations were calculated using the Spearman rank correlation. For regression modeling, LC-MS peak areas were log-transformed and scaled to a mean of 0 and SD of 1 within batch. For cross-sectional trait associations, logistic regression models adjusting for age and sex were used for the binary outcome of prevalent diabetes. Linear regression models were used for the continuous clinical traits of BMI and log-transformed FPG, HOMA-IR, and triglyceride levels due to values being right skewed. Cox proportional hazards models were used to calculate the hazard ratio (HR) and 95% CI for a 1-SD increase in compound concentration with incident diabetes. Three JHS models were defined a priori. Model 1 adjusted for age, sex, and batch to identify analytes associated with diabetes, including via increased adiposity and IR. Model 2 further adjusted for BMI and FPG, identifying associations that are independent of these two known biological mechanisms. Model 3 additionally adjusted for hypertension status, HDL cholesterol level, triglyceride levels, and statin use to identity diagnostic biomarkers that are independent of known diabetes risk factors.

Compounds with a CV >30% and/or >5% missingness in any batch were excluded; concentrations of those with <5% missingness were imputed at half of the lowest batch value. A Benjamini-Hochberg false discovery rate (FDR-q) <0.05 was used for significance to correct for the 1,434 compounds included in the analyses. A total of 263 compounds (including targeted and nontargeted) that were significantly associated with incident diabetes in JHS model 1 and 107 compounds from JHS model 2 were measured in MESA and were nominated for replication. Cox proportional hazards models that adjusted for *1*) age, sex, race/ethnicity, and batch and *2*) additionally for BMI and FPG were used with statistical significance defined at an FDR-q < 0.05.

To assess metabolite efficacy as diabetes predictive biomarkers, compounds were selected using elastic net regularization in Cox models for incident diabetes. The Harrell *c* statistic, Akaike information criteria, and Bayesian information criteria were calculated for models that included *1*) only clinical risk factors, *2*) risk factors and targeted compounds, *3*) risk factors and nontargeted compounds, and *4*) risk factors and both targeted and nontargeted compounds. The clinical risk factors included age, sex, BMI, SBP, HDL, triglycerides, waist circumference, FPG, and parental history of diabetes (31). Model discrimination was validated in MESA. The same prediction models were used except for the exclusion of parental history of diabetes due to data availability. All analyses were conducted using Stata and R statistical analysis software.

WGS Association Studies

WGS in JHS was obtained in participants who were included in Freeze 6 of the Trans-Omics for Precision Medicine (TOPMed) project at the University of Washington and Broad Institute; methods have been previously described (32).

Data and Resource Availability

The data sets generated during and/or analyzed during the current study were uploaded to the JHS database of Genotypes and Phenotypes (dbGaP) repository and/or are available upon request from the respective study cohorts, which can be facilitated by the corresponding author. MS/MS spectra of the unknown compounds were uploaded to the Global Natural Products Social Molecular Networking (GNPS) website under the job ID: aa6d11c8be15436abc b7d3d44fee5836. We also uploaded relevant MS/MS spectra, including those obtained from the Paternò-Büchi reaction under the Mass Spectrometry Interactive Virtual Environment (MassIVE) database, under data set MSV000090113 (doi:10.25345/C5V97ZW46), with a complete list of the spectra that were uploaded in the Supplementary Materials.

RESULTS

Baseline Characteristics

Baseline traits for the metabolomics subcohort (n = 2,750) are summarized in Table 1 and were similar to the whole JHS cohort (n = 5,306) (Supplementary Table 1). Of the 1,700 individuals in the incident analysis, the 315 cases were more likely to be older, had hypertension, and used statin medications. They also had higher BMI, FPG, and triglyceride levels and lower eGFR and HDL cholesterol.

Compound Feature Correlations

Spearman rank correlations were calculated between the targeted (i.e., known) and nontargeted (i.e., unknown) compounds measured using the hybrid HILIC-positive method (Fig. 1 heat map). Among known compounds, stronger correlations were seen among those from the same class. For example, compared with a median ρ = 0.55 among all knowns, valine had a median ρ = 0.93 with other branched chain amino acids. By contrast, unknown compounds were less correlated with each other (median ρ = 0.27) and knowns (median ρ = 0.29), suggesting that they may report on diverse metabolic processes.

Compound Associations With Prevalent Diabetes and Select Baseline Clinical Traits in JHS

There were 176 known compounds associated with diabetes at examination 1 after adjusting for age, sex, and batch (FDR-q < 0.05) (Fig. 2*A* and Supplementary Table 2).

Table 1—Baseline clinical characteristics of individuals included in the metabolomics analysis of the JHS

	Total (N = 2,750)	Control subjects (n = 1,385)	Case subjects (n = 315)
Age, years	55.9 (12.7)	53.5 (12.5)	55.9 (11.3)
Female, *n* (%)	1704 (62)	849 (61)	192 (61)
BMI, kg/m^2	31.7 (7.1)	30.5 (6.5)	33.9 (7.5)
FPG, mg/dL	101.9 (35.7)	89.2 (7.7)	97.0 (9.9)
SBP, mmHg	127.7 (16.7)	125.5 (15.6)	128.8 (17.5)
eGFR (CKD-EPI), mL/min/1.73 m^2	93.0 (22.5)	95.7 (20.2)	92.2 (19.1)
Total cholesterol, mg/dL	199.8 (41.0)	199.0 (38.6)	200.5 (42.3)
Triglycerides, mg/dL	109.0 (78.2)	95.6 (55.2)	117.2 (65.8)
HDL, mg/dL	51.8 (14.7)	53.1 (15.1)	49.0 (12.8)
Smoking, *n* (%)	335 (12)	156 (11)	37 (12)
Hypertension, *n* (%)	1,618 (59)	648 (47)	208 (66)
Statin use, *n* (%)	391 (14)	109 (8)	44 (14)
Baseline diabetes, *n* (%)	710 (26)	—	—

Values represent mean (SD) unless otherwise noted. Mean follow-up time was 10.2 years for incident diabetes. CKD-EPI, Chronic Kidney Disease Epidemiology Collaboration.

These included the inverse association of 1,5-anhydrosorbitol/1,5-anhydroglucitol, which is used clinically to measure hyperglycemic excursions (33) and has been nominated as a possible biomarker of sodium–glucose cotransporter inhibitor treatment efficacy (34). An additional 625 unknowns were also associated. Of these 801 total compounds, 269 remained significant after further adjustments for oral diabetes medication and/or insulin use (Supplementary Table 3).

There were 177 knowns and 535 unknowns associated with BMI, 156 knowns and 516 unknowns with FPG, 166 knowns and 509 unknowns with HOMA-IR, and 201 knowns and 710 unknowns with triglycerides. To quantify the percentage of variation in clinical trait explained by circulating compounds, R^2 values were calculated from least absolute shrinkage and selection operator regression models adjusted for age and sex that included known compounds alone and knowns and unknowns (Fig. 2*B*). The addition of nontargeted data increased the percentage of variance explained in all traits, dramatically so for FPG. While there was a significant number of overlapping compound associations with diabetes and these risk factors, several did not overlap, especially among the unknowns (Fig. 2*C* and *D* and Supplementary Table 4). For example, 55 known and 241 unknown compounds were associated with diabetes but not BMI, 65 knowns and 261 unknowns were associated with diabetes but not HOMA-IR, and 10 knowns and 47 unknowns were associated with prevalent diabetes alone.

Metabolite Associations With Incident Diabetes in JHS

There were 307 compounds, including 82 knowns, associated with diabetes incidence during a mean 10.2 years of follow-up in JHS model 1 (Supplementary Table 5), and 124 were associated in JHS model 2 (Fig. 3 and Supplementary Table 5). Thirty-five were known compounds, including previously reported associations such as the branched chain amino acids. Serine was associated with the lowest HR (HR 0.75 [95% CI 0.67–0.83], $q = 9.09 \times 10^{-5}$) and urate with the highest (HR 1.37 [95% CI 1.20–1.56], $q = 3.26 \times 10^{-4}$). Of these, 11 have not previously been reported in other human cohorts, and an additional 16 have not been found in cohorts that include AAs (Supplementary Table 6), including serotonin (HR 0.85, $q = 2.62 \times 10^{-2}$), homoarginine (HR 1.26, $q = 9.61 \times 10^{-3}$), and *N*-palmitoyl taurine (HR 1.29, $q = 2.26 \times 10^{-2}$). After further adjusting for hypertension status, statin medication use, and HDL and triglyceride levels, six metabolites remained significant (Supplementary Table 7). Of the 89 unknown compounds associated in model 2 (Supplementary Table 5), 19 remained after further adjustments for hypertension status, statin medication use, and lipid measurements (Supplementary Table 8).

Validation of Metabolite-Diabetes Associations in MESA

The incident diabetes associations were validated in 918 MESA participants, of whom 175 (19%) were self-reported AAs (Supplementary Table 9). Compared with JHS, MESA participants were older, had lower BMI and eGFR, had FPG that was higher in case subjects but lower in control subjects, and fewer were women.

Of 82 known compounds associated with diabetes in JHS model 1, 46 replicated in MESA model 1 (FDR-$q < 0.05$) (Supplementary Table 10) and 98 of 225 unknowns replicated. Of the 35 knowns associated with incident disease in JHS model 2, 25 replicated in MESA model 1 and 5 in MESA model 2 (FDR-$q < 0.05$) (Table 2). Of these five compounds, the inverse association of the plasmalogen lipid species phosphatidylethanolamine (PE)(P-36:2)/PE(O-36:3) (HR 0.68 [95% CI 0.56–0.83], $q = 6.42 \times 10^{-3}$) and phosphatidylcholine (PC)(P-34:2)/PC(O-34:3) (HR 0.71

Figure 1— Correlation of both targeted or known and nontargeted or unknown compounds measured using the HILIC-positive LC-MS method. Heat map of the Spearman rank correlation coefficients for baseline LC-MS compound concentrations in all JHS participants who underwent metabolomics profiling. The compounds included were measured using the targeted and nontargeted HILIC-positive LC-MS method that measured known and unknown compounds. Of note, only primary LC-MS features that are believed to represent unique compounds and not daughter ions, adducts, etc., are included. Strong positive correlations are dark red and strong negative correlations are dark blue. The grey bars denote targeted features and the green bars denote nontargeted features. Targeted feature correlations are in the left upper corner, targeted to nontargeted feature correlations are in the right upper corner, and nontargeted feature correlations are in the lower right.

[95% CI 0.58–9.87], q = 0.02) has not previously been reported. Of the 89 unknown compounds from JHS model 2, 43 validated in MESA model 1 (Supplementary Table 10) and 3 validated in MESA model 2 (Table 2).

Integration of High Mass Accuracy Spectrometry and Human Genetics Identifies a Novel Biomarker of Diabetes

Compound QI15902, with an RT of 1.72 min and m/z of 634.6486, had an HR of 1.46 for incident diabetes (95% CI 1.29–1.66, $q = 1.67 \times 10^{-7}$) (Supplementary Table 5) in JHS model 1 and an HR 1.31 (95% CI 1.16–1.49, $q = 1.31 \times 10^{-3}$) in model 2. This replicated in MESA model 1 (HR 1.44 [95% CI 1.17–1.76], $q = 1.74 \times 10^{-3}$). QI15902 was clustered with four other nontargeted LC-MS peaks, several of which were even more strongly associated with diabetes (Supplementary Fig. 1 and Supplementary Table 11). In WGS, four of the five peaks were associated with the same genetic variant in the *MEIS2* gene on chromosome 15, rs1357470, three at GWS (QI15902, QI15886, and QI299,

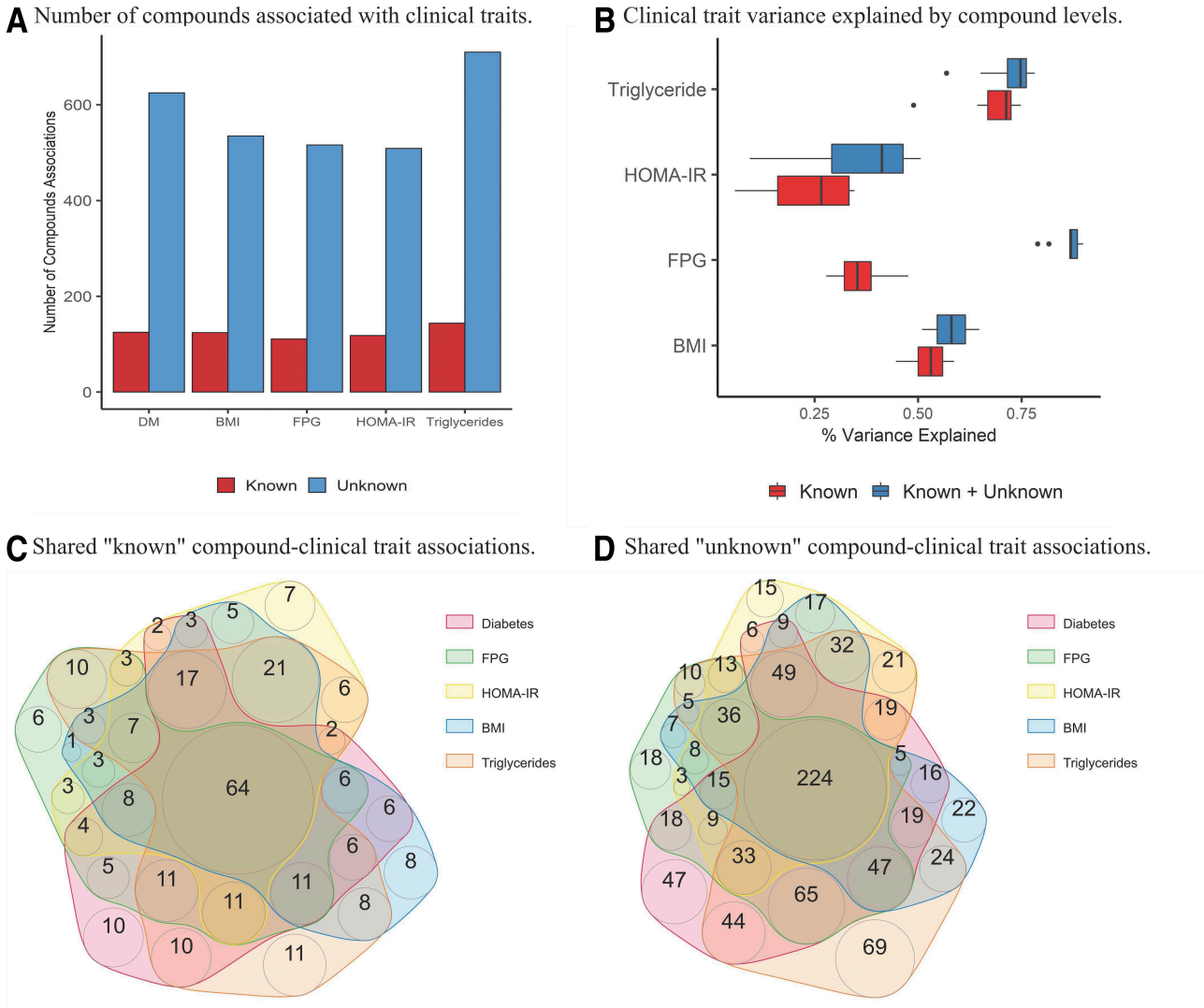


Figure 2—Known and unknown compound associations with clinical traits and prevalent diabetes. *A*: Bar graph represents the number of targeted (i.e., known) and nontargeted (i.e., unknown) compounds associated with select clinical traits. Clinical traits are plotted on the *x*-axis. The number of compounds associated with an FDR-$q < 0.05$ with the dichotomous outcome of prevalent diabetes in logistic regression models and with select continuous clinical traits in linear regression models are shown on the *y*-axis. All models were adjusted for age and sex. DM, diabetes mellitus. *B*: Percentage of clinical trait variability explained by known and unknown circulating compound levels. R^2 values, plotted on the *x*-axis, from adapted least absolute shrinkage and selection operator regression models that model the association of *1*) known and *2*) known and unknown compounds with BMI, FPG, HOMA-IR, and triglycerides are shown. These were restricted to the compounds measured using the hybrid HILIC positive method. Venn diagram shows the number of shared targeted or known compounds (*C*) and nontargeted or unknown compound associations (*D*) with prevalent diabetes, BMI, FPG, HOMA-IR, and clinical triglyceride levels.

with $P < 4.48 \times 10^{-8}$), and one at sub-GWS (QI15901 $P = 2.36 \times 10^{-7}$), supporting that these features were appropriately clustered. Two peaks (QI15902 and QI15886) were associated with a variant in the *CPS1* gene on chromosome 2 (rs1047891, $P < 4.1 \times 10^{-8}$). Variants in this gene have previously been associated (35,36)—and in the JHS were also associated—with circulating glycine and serine levels (35,36). The RT of QI15902 was consistent with a lipid species. Given the CPS1 polymorphism association with serine and glycine, which are participants in de novo ceramide synthesis, we postulated that this novel compound could be a lipid product of ceramide biosynthesis.

The parent ion mass (*m/z* 634.6486) and subppm MS/MS data collected on QI15902 matched a deoxyceramide, *N*-(tetracosanoyl)-1-deoxysphing-4-enine (Cer[m18:1/24:0]) in the CSI:FingerID database. LC-MS analysis of a synthetic Cer(m18:1/24:0) reference compound containing linear alky chains yielded MS and MS/MS spectra that matched QI15902 (Fig. 4); however, the RT of QI15902 did not match, suggesting it could be an isobaric species of Cer(m18:1/24:0) with differences in either the double-bond position or *cis*-orientation, or alkyl chain branching. To determine whether the double-bond position in QI15902 differed from Cer(m18:1/24:0), a plasma

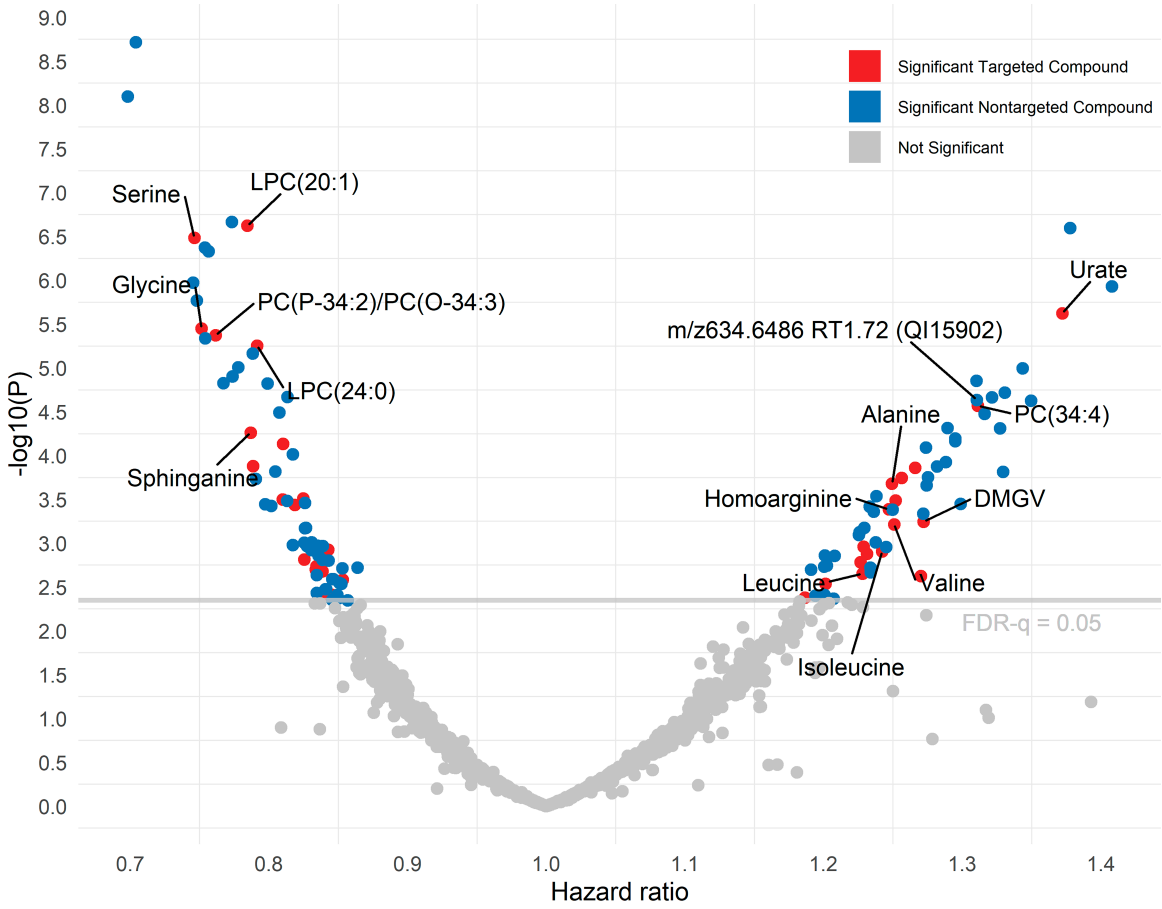


Figure 3—Known and unknown compound associations with incident diabetes in the JHS. Volcano plot shows targeted (i.e., known) and nontargeted (i.e., unknown) LC-MS compounds associated with incident diabetes in JHS model 2 (Cox proportional hazard model adjusted for age, sex, BMI, and FPG). Diabetes HRs over a mean 10.2 years of follow-up for every 1-SD increase in transformed and normalized LC-MS feature concentration are plotted on the *x*-axis. The colored dots represent metabolite features associated with diabetes at an FDR-$q < 0.05$. Red dots are known compounds, blue are unknown compounds. Select targeted LC-MS features are named. DMGV, dimethylguanidino valeric acid; LPC, lipophopsphatyidlycholine; TG, triacylglycerol.

sample was fractionated using C8 chromatography, and photochemical Paternò-Büchi reaction (37) double-bond cleavage products were generated from the fraction containing QI15902 (Supplementary Materials and Supplementary Fig. 2). LC-MS analyses showed that the QI15902 major cleavage product was an ion at *m/z* 454 and was the same mass as that generated from the Cer(m18:1/24:0) reference compound. This indicated that QI15902 has a 4,5-double bond like Cer(m18:1/24:0). This cleavage product also contained the C24 fatty acid moiety. Notably, cleavage of QI15902 yielded at least two different isomeric peaks, the more abundant of which had an earlier RT compared with the product from Cer(m18:1/24:0). These data indicate that QI15902 has a sphingoid base similar to Cer(m18:1/24:0), but the structure of the C24 fatty acid is different. The current hypothesis is that QI15902 is an isomer of Cer(m18:1/24:0) possessing a branched alkyl group in the C24 fatty acid moiety of the molecule. Predicted compound identities for other nontargeted features associated with incident diabetes are listed in Supplementary Table 12, and definitive identification using commercially available chemical standards are ongoing.

Utility of Circulating Compounds in Diabetes Prediction

A clinical diabetes risk prediction model based on the Framingham diabetes risk score (including age, sex, BMI, SBP, HDL, triglycerides, waist circumference, FPG, and parental history of diabetes) had a *c* statistic of 0.74 in JHS (Fig. 5). This improved to 0.77 with the addition of 10 known compounds selected using elastic net regularization. A similar improvement occurred with the addition of 11 unknown compounds. Inclusion of both knowns ($n = 9$) and unknowns ($n = 26$) further improved the *c* statistic to 0.81 (Fig. 5). Incident receiver operating characteristic curves over the course of 10 years of the different models are shown in Supplementary Fig. 3. Reclassification of case subjects and control subjects to high- and low-risk groups calculated using the net reclassification index were also improved, especially with the inclusion of nontargeted

Table 2—Known and unknown compounds associated with incident diabetes in JHS that validated in MESA

Study compound name	JHS ID	*m/z*	RT	HMDB ID	Compound name	Adduct	Primary feature	Targeted	JHS model 2						MESA model 2					
									HR	95% CI	*P*	FDR	*n*	MESA ID	*m/z*	RT	HR	95% CI	FDR *q*	*n*
QI11485	QI11485	714.5421	5.62	NA	PE(P-35:2)/PE(O-35:3)	NA	True	0	0.83	0.74–0.93	1.79E−03	2.46E−02	1,672	QI9400	714.545	5.44	0.68	0.57–0.82	5.24E−03	904
C36:3 PE plasmalogen	QI4677	728.5579	5.6	HMDB0011441	PE(P-36:2)/PE(O-36:3)	NA	True	1	0.84	0.75–0.94	2.73E−03	3.29E−02	1,672	QI102	728.561	5.42	0.68	0.56–0.83	7.80E−03	904
Isoleucine	TF22	132.102	7.29	HMDB0000172	Isoleucine	[M+H]+	True	1	1.28	1.12–1.47	3.28E−04	8.32E−03	1,672	TF47	132.102	7.08	1.50	1.19–1.88	1.82E−02	904
Leucine	TF24	132.102	7.16	HMDB0000687	Leucine	[M+H]+	True	1	1.25	1.09–1.43	1.41E−03	2.06E−02	1,672	TF48	132.102	6.95	1.49	1.18–87	1.82E−02	904
C34:3 PC plasmalogen	QI10124	742.5736	6.75	HMDB0011211	PC(P-34:2)/PC(O-34:3)	[M+H]+	True	1	0.75	0.66–0.84	2.71E−06	3.51E−04	1,672	QI94	742.576	6.6	0.71	0.58–0.87	1.92E−02	904
QI9685	QI9685	440.1716	7.62	HMDB0000883	Valine	[M+UNK]+	True	0	1.34	1.17–1.54	2.52E−05	1.88E−03	1,672	QI7595	440.173	7.44	1.49	1.18–1.9	1.92E−02	904
Glycine	QI9373	76.0393	8.09	HMDB0000123	Glycine	[M+H]+	True	1	0.74	0.65–0.84	3.06E−06	3.71E−04	1,672	TF41	76.0395	7.86	0.71	0.57–0.88	3.21E−02	904
QI4952	QI4952	583.2547	5.79	NA	NA	NA	True	0	0.84	0.77–0.9)	1.48E−04	4.63E-03	1,672	QI8786	583.257	5.64	1.31	1.1–1.56	3.21E−02	902

Metabolites listed are both targeted or "known" and nontargeted or "unknown" LC-MS compounds that were associated with incident diabetes in JHS model 2 (Cox proportional hazard models adjusted for age, sex, metabolomics batch, BMI, and FPG) at an FDR $q < 0.05$ and that replicated in MESA model 2 (adjusted for age, sex, self-reported race/ethnicity, BMI, and FPG) at an FDR $q < 0.05$. Nontargeted features included for validation were those designated as the primary feature of their metabolite cluster. HMDB, Human Metabolome Database; NA, not applicable.

compounds compared with the clinical model alone (Supplementary Table 13). These prediction models were externally validated in 500 individuals from MESA, with an improvement in model discrimination again observed with the inclusion of both the knowns and unknowns (P = 0.009) (Supplementary Table 14).

DISCUSSION

We identified novel circulating compounds associated with incident diabetes in a large AA cohort and validated these associations in an independent multiethnic cohort. We expanded the number of known and unknown compounds measured by leveraging a hybrid targeted and nontargeted LC-MS method. Compared with knowns, unknown compounds are less correlated (median ρ = 0.29) and improved diabetes prediction model discrimination (*c* statistic increase from 0.77 to 0.81), suggesting they provide additional, orthogonal information. Finally, we combined functional genomic analyses with high resolution and accurate MS to identify a novel deoxyceramide biomarker of incident disease with one of the highest diabetes HRs found in JHS.

We identified 124 circulating compounds associated with incident diabetes after adjusting for age, sex, batch, BMI, and FPG. Of these, 35 were known compounds, 11 of which had not previously been reported in other human populations, and an additional 16 had not been found in AA cohorts (Supplementary Table 6). Sixty-eight of these associations replicated in MESA (FDR-q < 0.05 after adjusting for age, sex, batch, and race/ethnicity). Serotonin is a neuroactive amino acid known to contribute to glucose homeostasis (38) and was inversely associated with incident diabetes in our cohort. Homoarginine is a substrate for nitric oxide synthase (39) and was found to be significantly higher in AAs compared with Whites in the Dallas Heart Study (DHS) and positively associated with obesity, IR, and dysglycemia, but inversely associated with diabetes prevalence (40). We demonstrate a positive association that replicated in MESA. The fatty acid conjugated amino acid, *N*-palmitoyl taurine, accumulates in human islet cells and may be an insulin secretagogue (41), supporting our positive associations with prevalent and incident diabetes, BMI, and HOMA-IR.

Nontargeted metabolomics remains relatively unexplored because peak acquisition, data cleaning, and compound identification remain labor and time intensive. We demonstrate that nontargeted or unknown compounds are only modestly associated with each other (median ρ = 0.27) and targeted or known compounds (median ρ = 0.29, with median ρ = 0.55 among knowns). Inclusion of unknowns at least tripled the number of cross-sectional associations found with clinical traits (Fig. 2*A*). Unknowns also explained up to 45% of clinical trait variance (Fig. 2*B*). Several unknowns associated with diabetes were also not associated with traditional risk factors (Fig. 2*C* and *D*), providing potential insights into previously unknown pathways of disease development. For

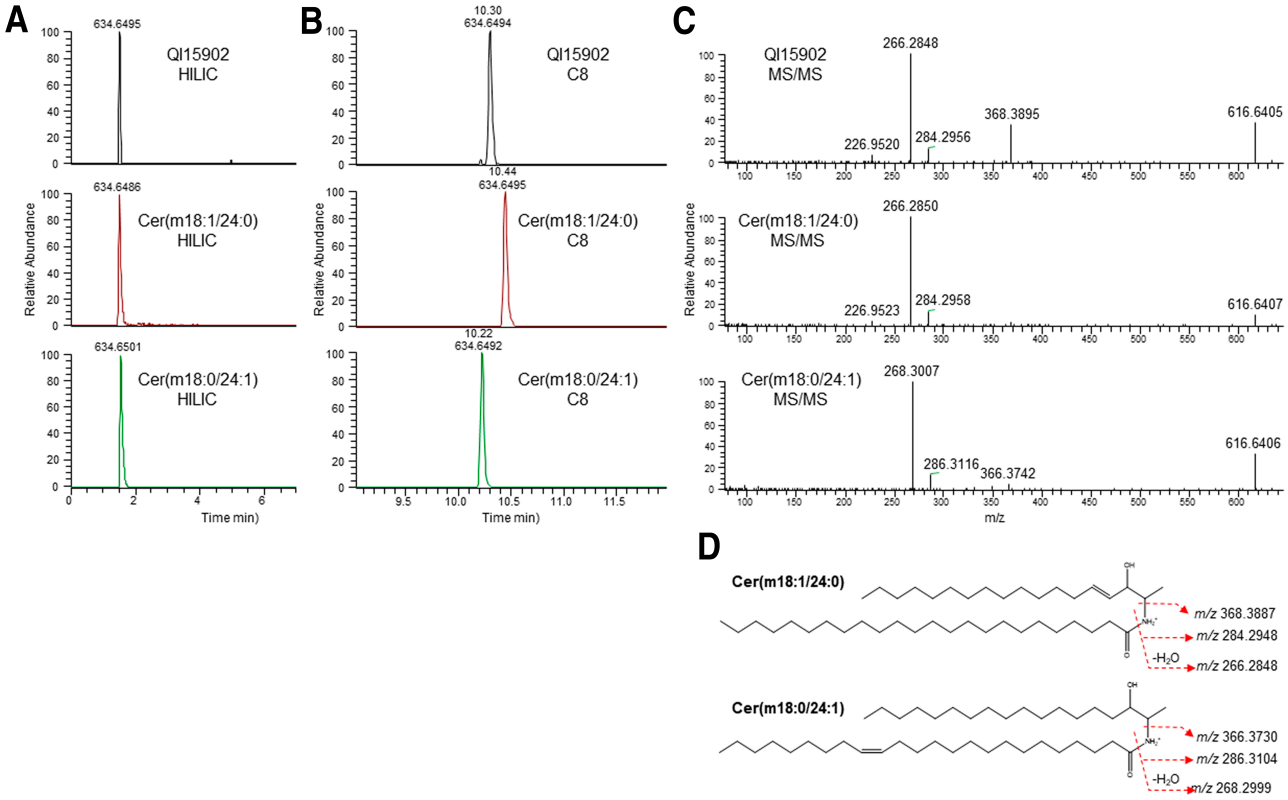
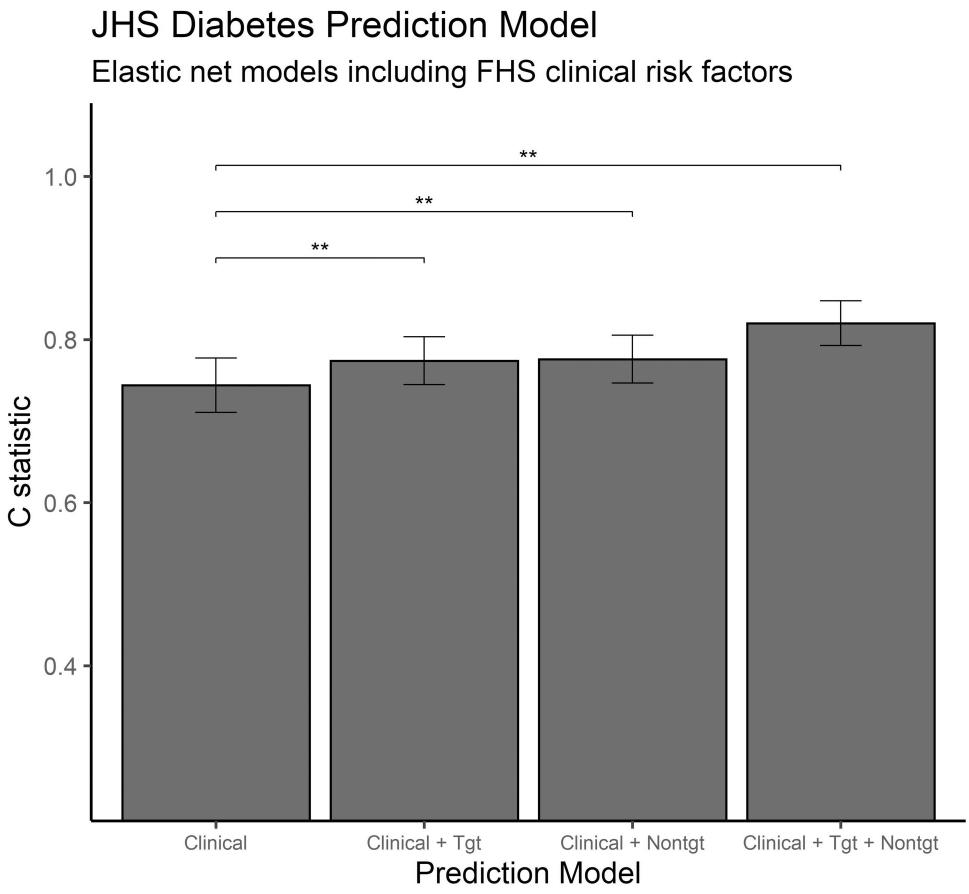


Figure 4—LC-MS and MS/MS of putative deoxyceramide. Extracted ion chromatograms for plasma unknown feature QI15902 in plasma compared to neat standards for Cer(m18:1/24:0) and Cer(m18:0/24:1) resolved using (*A*) hydrophilic liquid interaction chromatography (HILIC) or (*B*) reversed phase C8 chromatography. (*C*) Product ion spectra (MS/MS) for plasma unknown feature QI15902 compared to Cer(m18:1/24:0) and Cer(m18:1/24:0) standards. (*D*) Structures and MS/MS fragmentation of deoxyceramide standards. Extracted ion chromatograms of m/z 634.6496 ± 5 ppm and MS/MS spectra were obtained from the molecular ions ([M+H]+) at m/z 634.65 with an isolation window of 0.4 a.m.u.

example, compounds associated with diabetes, but not BMI and HOMA-IR, may participate in metabolic processes that cause diabetes independent of adiposity and IR.

In JHS, 225 unknown compounds were associated with incident diabetes after adjusting for age and sex, and 89 were associated with additional adjustments for BMI and FPG (108 and 43 validated in MESA, respectively). Included in these associations was a novel isobaric species of 1-deoxyceramide Cer(m18:1/24:0) that we identified by leveraging WGS data. This novel compound, or QI15902, had an HR of 1.38, similar to urate (HR 1.37), which was the highest among knowns. QI15902 was inversely associated with a *CPS1* gene variant (rs1047891, $\beta = -0.17$, $P = 1.48 \times 10^{-8}$, minor allele frequency = 0.36). *CPS1* encodes for carbamoyl-phosphate synthase 1, which catalyzes the first committed step of the urea cycle, and this variant has been associated with circulating glycine and serine levels (35,36), including in the JHS ($\beta = 0.20$, $P = 5.03 \times 10^{-11}$ for serine). De novo synthesis of sphingolipids and ceramides are initiated by the condensation of serine and palmitoyl CoA. QI15902 had a RT suggestive of a lipid species. In the absence of serine, alanine is condensed with palmitoyl CoA to form deoxyceramides and deoxysphinganine (42). Given the inverse association of QI15902 with rs1047891, we hypothesized it was part of the deoxyceramide pathway. Consistent with this, individuals in the JHS with the 4217C>A missense gene variation also had higher levels of circulating serine and glycine, and lower levels of alanine and QI15902 (Supplementary Fig. 4). Finally, after further MS work, we have confirmed that QI15902 is an isobaric 1 deoxyceramide Cer(m18:1/24:0) species.

Elevated levels of 1-deoxysphingolipid and 1-deoxysphinganine—which are closely related to deoxyceramides—are found among individuals with metabolic syndrome (43,44), impaired fasting glucose (44), impaired oral glucose tolerance in pregnant women (45), and diabetes (44,46). Owing to a missing hydroxyl group, these complex lipids cannot be degraded, leading to cellular accumulation and possible toxicity (47). Deoxyceramides, specifically, are positively associated with neuropathy in individuals with type 1 diabetes (48), but its association with T2D, especially in AA cohorts, has not been extensively studied. Interestingly, QI15902 along with three other LC-MS peaks from this compound cluster were also associated with a variant in the developmental gene *MEIS2* (49), and further studies are needed to determine whether *MEIS2* may serve as a master regulator of *CPS1*. Several of the novel known compound associations with incident diabetes



Model	C-statistic	AIC	BIC
Clinical	0.735	3249	3296
Clinical + Targeted LC-MS Compounds	0.769	3216	3321
Clinical + Nontargeted LC-MS Compounds	0.766	3219	3323
Clinical + Targeted + Nontargeted LC-MS Compounds	0.811	3160	3390

Figure 5—Targeted and nontargeted metabolite diabetes risk discrimination. The *c* statistics are shown for prediction models of incident diabetes in JHS. Clinical risk factors included age, sex, BMI, SBP, HDL cholesterol, triglycerides, FPG, and family history of diabetes and were used in all models. Elastic net regularization was used to select targeted (tgt) and nontargeted (nontgt) LC-MS primary features to be included as predictors in the respective models. **Denotes likelihood ratio test. $P < 0.001$ comparing nested models.

in JHS were also lipid subspecies (Supplementary Table 10). Replication of these associations in MESA, however, were varied. Whether these differences are due to diet or heterogeneity in genetic makeup across these cohorts is an important question to answer and motivates dedicated lipidomic profiling to improve measurement specificity and in-depth genetic association studies to further explore.

Finally, in clinical prediction models, the addition of both known and unknown compounds improved model discrimination in a stepwise fashion in JHS (Fig. 5). There was a significant increase in the model *c* statistic and an AIC that favored the use of a combined clinical, known, and unknown compound prediction model. A modest increase in the *c* statistic was also observed with the inclusion of metabolite predictors in the multiethnic MESA cohort (Supplementary Table 12); however, a limitation was the lack of family diabetes history data, which improves clinical prediction models. While the inclusion of these biomarkers may not be practical for the clinical diagnosis of diabetes, these models demonstrate that unknown circulating metabolites provide insights into diabetes beyond what is provided by knowns. Furthermore, as metabolites, these unknown compounds can highlight pathways that may contribute to disease development and complications that are both dependent and independent of dysglycemia, obesity, and insulin resistance and warrant further study.

Our study has many strengths, including the breadth and depth of our metabolomics profiling in a large cohort of AAs.

Limitations include the small number of AAs who had metabolomics profiling available in MESA; therefore, we were unable to replicate in a cohort with similar race/ethnicity makeup. While we found novel compound associations in a large cohort of self-reported AAs, we will need metabolomics and genetic data from other large multiethnic cohorts before we can draw conclusions about how self-reported race/ethnicity, genetic ancestry, and social determinants of health contribute to these associations and is a planned future direction of study. For our diabetes case definition, FPG and HbA_{1c} were used, but we did not have oral glucose tolerance tests, which could have led to case misclassification of some individuals. Also, while the majority of prevalent and incident cases were likely of T2D, C-peptide and islet autoantibodies were not measured so we could not exclude individuals who had type 1 diabetes.

In conclusion, using targeted and nontargeted LC-MS methods, we have identified novel incident diabetes metabolites in a population of self-reported AAs, with a majority that replicated in a multiethnic cohort. We identified a novel lipid species as a new biomarker of diabetes that warrants further mechanistic studies. Future steps will be to validate our findings in both multiethnic and ethnic-specific cohorts to understand how race, ethnicity, and social determinants of health may affect these metabolite-disease associations. Finally, we demonstrate that unknown metabolites provide added knowledge, explaining a significant amount of the variance in clinical traits associated with diabetes risk and prevalent and incident disease and improves clinical diabetes prediction model discrimination. These results motivate further studies focused on the identification of nontargeted LC-MS peaks to increase our understanding of diabetes biomarkers in diverse human populations.

Acknowledgments. The authors thank all the study participants and staff in the JHS and MESA for their important contributions.

Funding. Research in this manuscript was supported by the National Institute of Diabetes and Digestive and Kidney Diseases with K23DK127073 to Z.-Z.C. and R01DK081572 to R.E.G. and J.G.W. The JHS is supported by contracts from the National Heart, Lung, and Blood Institute and the National Institute for Minority Health and Health Disparities and conducted in collaboration with Jackson State University (HHSN268201800013I), Tougaloo College (HHSN268201800014I), the Mississippi State Department of Health (HHSN268201800015I/HHSN26800001), and the University of Mississippi Medical Center (HHSN268201800010I, HHSN268201800011I, and HHSN268201800012I). MESA and the MESA SNP Health Association Resource (SHARe) projects are conducted and supported by the National Heart, Lung, and Blood Institute in collaboration with MESA investigators. Support for MESA is provided by National Heart, Lung, and Blood Institute contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, and N01-HC-95169, and National Center for Advancing Translational Sciences grants UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420. WGS for the TOPMed program was supported by the National Heart, Lung and Blood Institute. Centralized read mapping and genotype calling, along with variant quality metrics and filtering, were provided by the TOPMed Informatics Research Center support from National Heart, Lung and Blood Institute (3R01HL117626-02S1, contract HHSN268201800002I) (Broad RNA sequencing, Proteomics HHSN268201600034I, University of Washington RNA sequencing HHSN268201600032I, University of Southern California DNA Methylation HHSN268201600034I, and Broad Metabolomics HHSN268201600038I). Phenotype harmonization, data management, sample-identity QC, and general study coordination were provided by the TOPMed Data Coordinating Center support from National Heart, Lung and Blood Institute (3R01HL-120393, U01HL-120393, contract HHSN268180001I). The provision of genotyping data were supported in part by the National Center for Advancing Translational Sciences, Clinical and Translational Science Institute grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center grant DK063491 to the Southern California Diabetes Endocrinology Research Center. Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute grant R01HL105756 and also by R01HL151855-01 and the National Institute of Diabetes and Digestive and Kidney Diseases UM1DK078616 contracts.

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute, the National Institutes of Health, or the U.S. Department of Health and Human Services.

Duality of Interest. No potential conflicts of interest relevant to this article were reported.

Author Contributions. Z.-Z.C. designed the study, led the data analysis, and drafted the manuscript. J.A.P. performed most of the LC-MS and MS/MS profiling that facilitated chemical identification of the unknowns. J.A.P., Y.G., S.D., B.P., and X.G. performed the statistical analyses and contributed to the drafting of the manuscript. X.S. and S.Z. helped perform the LC-MS compound profiling of the cohorts. U.A.T., D.H.K., D.E.C., M.D.B., and J.M.R. conducted/contributed to the analysis of the WGS analyses in JHS. D.N., J.G.W., C.B.C., and R.E.G. provided mentorship and critical feedback at all stages of the project, including during the drafting of the manuscript. M.d.R.S.G., A.M., L.A.L., A.C., M.J., K.D.T., S.S.R., M.O.G., and J.I.R. provided collaborative support within the different studies (JHS, MESA, and TOPMed) and reviewed and provided critical feedback for the manuscript. R.E.G. 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 and accuracy of the data and data analysis.

References

1. Centers for Disease Control and Prevention. National Diabetes Statistics Report, 2020. Accessed 16 March 2021. Available from https://www.cdc.gov/diabetes/data/statistics-report/index.html

2. Lanting LC, Joung IMA, Mackenbach JP, Lamberts SWJ, Bootsma AH. Ethnic differences in mortality, end-stage complications, and quality of care among diabetic patients: a review. Diabetes Care 2005;28:2280–2288

3. Hostalek U. Global epidemiology of prediabetes—present and future perspectives. Clin Diabetes Endocrinol 2019;5:5

4. Zhu Y, Sidell MA, Arterburn D, et al. Racial/ethnic disparities in the prevalence of diabetes and prediabetes by BMI: Patient Outcomes Research To Advance Learning (PORTAL) multisite cohort of adults in the U.S. Diabetes Care 2019;42:2211–2219

5. Unger RH, Orci L. Diseases of liporegulation: new perspective on obesity and related disorders. FASEB J 2001;15:312–321

6. Kahn HS, Cheng YJ, Thompson TJ, Imperatore G, Gregg EW. Two risk-scoring systems for predicting incident diabetes mellitus in U.S. adults age 45 to 64 years. Ann Intern Med 2009;150:741–751

7. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 2007;445:881–885

8. Zeggini E, Weedon MN, Lindgren CM, et al. Multiple type 2 diabetes susceptibility genes following genome-wide association scan in UK samples. Science 2007;316:1336–1341

9. Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. Nature 2016;536:41–47
10. Mahajan A, Spracklen CN, Zhang W, et al.; FinnGen; eMERGE Consortium. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. Nat Genet 2022;54:560–572
11. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. Nat Med 2011;17:448–453
12. Rhee EP, Cheng S, Larson MG, et al. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. J Clin Invest 2011;121:1402–1411
13. Wang-Sattler R, Yu Z, Herder C, et al. Novel biomarkers for pre-diabetes identified by metabolomics. Mol Syst Biol 2012;8:615
14. Stancáková A, Civelek M, Saleem NK, et al. Hyperglycemia and a common variant of GCKR are associated with the levels of eight amino acids in 9,369 Finnish men. Diabetes 2012;61:1895–1902
15. Cheng S, Rhee EP, Larson MG, et al. Metabolite profiling identifies pathways associated with metabolic risk in humans. Circulation 2012;125:2222–2231
16. Floegel A, Stefan N, Yu Z, et al. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. Diabetes 2013;62:639–648
17. Ferrannini E, Natali A, Camastra S, et al. Early metabolic markers of the development of dysglycemia and type 2 diabetes and their physiological significance. Diabetes 2013;62:1730–1737
18. Padberg I, Peter E, González-Maldonado S, et al. A new metabolomic signature in type-2 diabetes mellitus and its pathophysiology. PLoS One 2014;9:e85082
19. Palmer ND, Stevens RD, Antinozzi PA, et al. Metabolomic profile associated with insulin resistance and conversion to diabetes in the Insulin Resistance Atherosclerosis Study. J Clin Endocrinol Metab 2015;100:E463–E468
20. Rebholz CM, Yu B, Zheng Z, et al. Serum metabolomic profile of incident diabetes. Diabetologia 2018;61:1046–1054
21. Psychogios N, Hau DD, Peng J, et al. The human serum metabolome. PLoS One 2011;6:e16957
22. Sempos CT, Bild DE, Manolio TA. Overview of the Jackson Heart Study: a study of cardiovascular diseases in African American men and women. Am J Med Sci 1999;317:142–146
23. Bild DE, Bluemke DA, Burke GL, et al. Multi-Ethnic Study of Atherosclerosis: objectives and design. Am J Epidemiol 2002;156:871–881
24. O'Sullivan JF, Morningstar JE, Yang Q, et al. Dimethylguanidino valeric acid is a marker of liver fat and predicts diabetes. J Clin Invest 2017;127:4394–4402
25. Carpenter MA, Crow R, Steffes M, et al. Laboratory, reading center, and coordinating center data management methods in the Jackson Heart Study. Am J Med Sci 2004;328:131–144
26. Levey AS, Stevens LA, Schmid CH, et al.; CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate. Ann Intern Med 2009;150:604–612
27. Bertoni AG, Kramer H, Watson K, Post WS. Diabetes and clinical and subclinical CVD. Glob Heart 2016;11:337–342
28. Kimberly WT, O'Sullivan JF, Nath AK, et al. Metabolite profiling identifies anandamide as a biomarker of nonalcoholic steatohepatitis. JCI Insight 2017;2:92989
29. Paynter NP, Balasubramanian R, Giulianini F, et al. Metabolic predictors of incident coronary heart disease in women. Circulation 2018;137:841–853
30. Dührkop K, Fleischauer M, Ludwig M, et al. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. Nat Methods 2019;16:299–302
31. Wilson PWF, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB Sr. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. Arch Intern Med 2007;167:1068–1074
32. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature 2021;590:290–299
33. McGill JB, Cole TG, Nowatzke W, et al.; U.S. trial of the GlycoMark assay. Circulating 1,5-anhydroglucitol levels in adult patients with diabetes reflect longitudinal changes of glycemia: a U.S. trial of the GlycoMark assay. Diabetes Care 2004;27:1859–1865
34. Usui M, Tanaka M, Takahashi H. 1,5-anhydroglucitol is a good predictor for the treatment effect of the sodium-glucose cotransporter 2 inhibitor in Japanese patients with type 2 diabetes mellitus. J Clin Transl Endocrinol 2020;21:100233
35. Shin SY, Fauman EB, Petersen AK, et al.; Multiple Tissue Human Expression Resource (MuTHER) Consortium. An atlas of genetic influences on human blood metabolites. Nat Genet 2014;46:543–550
36. Imaizumi A, Adachi Y, Kawaguchi T, et al. Genetic basis for plasma amino acid concentrations based on absolute quantification: a genome-wide association study in the Japanese population. Eur J Hum Genet 2019;27:621–630
37. Murphy RC, Okuno T, Johnson CA, Barkley RM. Determination of double bond positions in polyunsaturated fatty acids using the photochemical Paternò-Büchi reaction with acetone and tandem mass spectrometry. Anal Chem 2017;89:8545–8553
38. Sumara G, Sumara O, Kim JK, Karsenty G. Gut-derived serotonin is a multifunctional determinant to fasting adaptation. Cell Metab 2012;16:588–600
39. Bretscher LE, Li H, Poulos TL, Griffith OW. Structural characterization and kinetics of nitric-oxide synthase inhibition by novel N5-(iminoalkyl)- and N5-(iminoalkenyl)-ornithines. J Biol Chem 2003;278:46789–46797
40. Atzler D, Gore MO, Ayers CR, et al. Homoarginine and cardiovascular outcome in the population-based Dallas Heart Study. Arteriosclerosis, Thrombosis, and Vascular Biology. Arterioscler Thromb Vasc Biol 2014;34:2501–2507
41. Aichler M, Borgmann D, Krumsiek J, et al. N-acyl taurines and acylcarnitines cause an imbalance in insulin synthesis and secretion provoking β cell dysfunction in type 2 diabetes. Cell Metab 2017;25:1334–1347.e4
42. Zitomer NC, Mitchell T, Voss KA, et al. Ceramide synthase inhibition by fumonisin B1 causes accumulation of 1-deoxysphinganine: a novel category of bioactive 1-deoxysphingoid bases and 1-deoxydihydroceramides biosynthesized by mammalian cell lines and animals. J Biol Chem 2009;284:4786–4795
43. Othman A, Rütti MF, Ernst D, et al. Plasma deoxysphingolipids: a novel class of biomarkers for the metabolic syndrome? Diabetologia 2012;55:421–431
44. Othman A, Saely CH, Muendlein A, et al. Plasma 1-deoxysphingolipids are predictive biomarkers for type 2 diabetes mellitus. BMJ Open Diabetes Res Care 2015;3:e000073
45. Khan A, Hornemann T. Correlation of the plasma sphingoid base profile with results from oral glucose tolerance tests in gestational diabetes mellitus. EXCLI J 2017;16:497–509
46. Bertea M, Rütti MF, Othman A, et al. Deoxysphingoid bases as plasma markers in diabetes mellitus. Lipids Health Dis 2010;9:84
47. Zuellig RA, Hornemann T, Othman A, et al. Deoxysphingolipids, novel biomarkers for type 2 diabetes, are cytotoxic for insulin-producing cells. Diabetes 2014;63:1326–1339
48. Hammad SM, Baker NL, El Abiad JM, et al.; DCCT/EDIC Group of Investigators. Investigators DG of. Increased plasma levels of select deoxyceramide and ceramide species are associated with increased odds of diabetic neuropathy in type 1 diabetes: a pilot study. Neuromolecular Med 2017;19:46–56
49. Paige SL, Thomas S, Stoick-Cooper CL, et al. A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. Cell 2012;151:221–232