



# Novel Metabolic Markers for the Risk of Diabetes Development in American Indians

*Diabetes Care* 2015;38:220–227 | DOI: 10.2337/dc14-2033

Jinying Zhao,<sup>1</sup> Yun Zhu,<sup>1</sup> Noorie Hyun,<sup>2</sup>  
Donglin Zeng,<sup>2</sup> Karan Uppal,<sup>3</sup>  
ViLinh T. Tran,<sup>3</sup> Tianwei Yu,<sup>4</sup> Dean Jones,<sup>3</sup>  
Jiang He,<sup>1</sup> Elisa T. Lee,<sup>5</sup> and  
Barbara V. Howard<sup>6</sup>

## OBJECTIVE

To identify novel metabolic markers for diabetes development in American Indians.

## RESEARCH DESIGN AND METHODS

Using an untargeted high-resolution liquid chromatography–mass spectrometry, we conducted metabolomics analysis of study participants who developed incident diabetes ( $n = 133$ ) and those who did not ( $n = 298$ ) from 2,117 normoglycemic American Indians followed for an average of 5.5 years in the Strong Heart Family Study. Relative abundances of metabolites were quantified in baseline fasting plasma of all 431 participants. Prospective association of each metabolite with risk of developing type 2 diabetes (T2D) was examined using logistic regression adjusting for established diabetes risk factors.

## RESULTS

Seven metabolites (five known and two unknown) significantly predict the risk of T2D. Notably, one metabolite matching 2-hydroxybiphenyl was significantly associated with an increased risk of diabetes, whereas four metabolites matching PC (22:6/20:4), (3S)-7-hydroxy-2',3',4',5',8-pentamethoxyisoflavan, or tetrapeptides were significantly associated with decreased risk of diabetes. A multimarker score comprising all seven metabolites significantly improved risk prediction beyond established diabetes risk factors including BMI, fasting glucose, and insulin resistance.

## CONCLUSIONS

The findings suggest that these newly detected metabolites may represent novel prognostic markers of T2D in American Indians, a group suffering from a disproportionately high rate of T2D.

Type 2 diabetes (T2D) is a metabolic disorder characterized by hyperglycemia resulting from impaired insulin secretion and increased insulin resistance (1). The pathogenesis of T2D is complex, involving both genetic and environmental factors, but the precise mechanisms underlying T2D development remain incompletely understood. Traditional risk factors such as age, sex, obesity, fasting glucose, and insulin resistance contribute considerably to disease risk and have therefore been widely used for routine diagnosis or risk stratification, but most of these markers fail to capture the complexity of disease etiology and thus have limitations in detecting early metabolic abnormalities that may occur years or even decades before the onset/diagnosis of overt T2D. Characterization of metabolic profiles and perturbed

<sup>1</sup>Department of Epidemiology, Tulane University School of Public Health, New Orleans, LA

<sup>2</sup>Department of Biostatistics, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC

<sup>3</sup>Division of Pulmonary Medicine, Emory University School of Medicine, Atlanta, GA

<sup>4</sup>Department of Biostatistics and Bioinformatics, Emory University School of Public Health, Atlanta, GA

<sup>5</sup>Center for American Indian Health Research, University of Oklahoma Health Sciences Center, Oklahoma City, OK

<sup>6</sup>Medstar Health Research Institute and Georgetown and Howard Universities Centers for Translational Sciences, Washington, DC

Corresponding author: Jinying Zhao, [jzhao5@tulane.edu](mailto:jzhao5@tulane.edu).

Received 25 August 2014 and accepted 31 October 2014.

This article contains Supplementary Data online at <http://care.diabetesjournals.org/lookup/suppl/doi:10.2337/dc14-2033/-/DC1>.

© 2015 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered.

See accompanying articles, pp. 186, 189, 197, 206, 213, and 228.

metabolic pathways implicated in T2D development will not only provide novel insights into disease pathophysiology but also provide instrumental data for risk prediction and for developing effective therapeutic and preventive strategies against diabetes.

Metabolomics is an emerging analytical technology that simultaneously quantifies many metabolites in biofluids. These metabolites represent the end products of cellular metabolism in response to intrinsic and extrinsic stimuli and thus may reflect the metabolic changes at earlier stages of disease. Cross-sectional analyses have reported associations of altered metabolites with obesity (2), insulin resistance (3), prediabetes, and overt T2D (4–7). These changes included acylcarnitines (6,8), amino acids (2,8), sugars (5,7), and different lipid species (5,8,9). Higher plasma levels of branched-chain amino acids (BCAAs) and aromatic amino acids were associated with an increased risk of T2D in the Framingham Offspring study (10). Another study found that increased diacyl-phosphatidylcholines and reduced acyl-alkyl- and lyso-phosphatidylcholines as well as sphingomyelins were associated with diabetes in a European population (11). More recently,  $\alpha$ -hydroxybutyrate and linoleoylglycerophosphocholine were also found to predict the development of dysglycemia and T2D in Europeans (12). These findings derived from European populations, however, may not represent metabolic alterations in other ethnic groups. Moreover, most existing studies used a targeted metabolomics approach by focusing on a subset of preselected metabolites and thus may have limited ability in discovering novel disease-related metabolic changes. The clinical utility of previously detected metabolites in risk prediction was either not reported or was minimal over conventional clinical factors.

The goal of this study is to identify predictive metabolic markers for future risk of T2D in American Indians, a minority group suffering from a disproportionately high rate of T2D. Metabolic profiles of diabetes development were examined in normoglycemic participants using fasting plasma samples collected prior to disease occurrence. The utility of novel metabolic markers in risk prediction beyond established diabetes risk factors was also investigated.

## RESEARCH DESIGN AND METHODS

### Study Population

Participants included in the current study were selected from the Strong Heart Family Study (SHFS), a family-based prospective study designed to identify genetic factors for cardiovascular disease (CVD), diabetes, and their risk factors in American Indians residing in Arizona, North and South Dakota, and Oklahoma. A detailed description for the study design and methods of the SHFS had been reported previously (13,14). In brief, a total of 3,665 tribal members (aged 14 years and older) from 94 multiplex families (65 three-generation and 29 two-generation families, average family size 38) were recruited and examined in 2001–2003. All living participants were followed and reexamined between 2006 and 2009. The SHFS protocol was approved by the institutional review boards from the Indian Health Service and the participating study centers. All participants gave informed consent.

According to the American Diabetes Association 2003 criteria (15), diabetes was defined as fasting plasma glucose  $\geq 7.0$  mmol/L or hypoglycemic medications. Impaired fasting glucose was defined as a fasting glucose of 6.1–6.9 mmol/L and no hypoglycemic medications, and normal fasting glucose (NFG) was defined as fasting glucose  $< 6.1$  mmol/L. Incident cases of T2D were defined as normal fasting glucose at baseline (2001–2003) and development of new T2D by the end of follow-up (2006–2009).

Participants included in the current analysis have to meet the following criteria: 1) attended clinical examinations at both baseline (2001–2003) and follow-up (2006–2009), 2) had NFG at baseline, 3) were free of overt CVD and hypoglycemic medications at baseline, and 4) had available fasting plasma sample at baseline for the proposed metabolomic analysis. Participants with missing information for fasting glucose or antidiabetes medication at either baseline or follow-up were also excluded from the current analysis.

A total of 2,324 participants free of overt CVD at baseline attended both clinical visits and had available fasting plasma samples for the proposed analysis. Of these, 2,117 normoglycemic participants met all of the criteria listed

above. After an average 5.5 years of follow-up, 197 participants (9.3%) developed incident T2D. Among those who did not develop T2D ( $n = 1,920$ ), 159 participants (7.5%) progressed to impaired fasting glucose, whereas the other individuals ( $n = 1,761$ ) remained with stable NFG by the end of follow-up. The current metabolomics analysis measured metabolite levels in fasting plasma of 431 participants, including 133 incident cases randomly selected from participants who developed new T2D ( $n = 197$ ) and 298 control subjects randomly selected from those who did not develop T2D ( $n = 1,920$ ). Supplementary Table 1 shows the comparison of baseline clinical characteristics between participants who were selected and those not selected.

### Assessments of Diabetes Risk Factors

Fasting plasma glucose, insulin, lipids, lipoproteins, and inflammatory biomarkers were measured by standard laboratory methods (14,16). BMI was calculated as body weight in kilograms divided by the square of height in meters. Hypertension was defined as blood pressure levels  $\geq 140/90$  mmHg or use of antihypertensive medications. Insulin resistance was assessed using HOMA according to the following formula: HOMA of insulin resistance (HOMA-IR) = fasting glucose (mg/dL)  $\times$  insulin ( $\mu$ U/mL)/405 (17). Renal function was assessed using the estimated glomerular filtration rate (eGFR) calculated by the MDRD equation (18). For cigarette smoking, subjects were classified as current smokers, former smokers, and nonsmokers. Alcohol consumption was determined by self-reported history of alcohol intake, the type of alcoholic beverages consumed, frequency of alcohol consumption, and average quantity consumed per day and per week. Participants are classified as current drinkers, former drinkers, and never drinkers. Dietary intake was assessed using the block food frequency questionnaire (19).

### Metabolic Profiling by High-Resolution Liquid Chromatography–Mass Spectrometry

Relative abundance of fasting plasma metabolites was determined using high-resolution liquid chromatography–mass spectrometry (LC-MS). Detailed laboratory protocols have previously been described (20,21). Briefly, 65  $\mu$ L plasma

sample aliquots were treated with acetonitrile, spiked with internal standard mix, and centrifuged at 13,000g for 10 min at 4°C to remove proteins. Supernatant (130  $\mu$ L) was removed and loaded into autosampler vials. Anion exchange (AE) columns (both C18 and AE columns) were equilibrated to the initial condition for 1.5 min prior to the next sample injection. Mass spectral data were collected with a 10-min gradient on a Thermo LTQ-Velos Orbitrap mass spectrometer (Thermo Fisher, San Diego, CA) to collect data from mass/charge ratio ( $m/z$ ) 85–2,000 in a positive ionization mode. Three technical replicates were run for each sample using a dual-column chromatography procedure with C18 and an AE column. Pooled plasma samples were included in each batch ( $n = 23$ ) for quality control. Peak extraction, data alignment, and feature quantification were performed using the adaptive processing software (apLCMS) (22,23), a computer package designed for high-resolution metabolomics data analysis. Feature and sample quality assessment was performed based on coefficient of variation (CV) and Pearson correlation, respectively, based on the technical replicates using xMSanalyzer (24). Metabolites with CV >50% in our samples were excluded from further analyses. Potential metabolite identities were determined by performing an online search (10 ppm mass accuracy) against the Metlin database (25), the Human Metabolomics Database (26), and the LIPID MAPS structure database (27). Data filtering, normalization, diagnostics, and summarization were performed using the computer package MSPrep (28). Missing data were imputed using the half of the minimum observed value within each metabolite across all samples. Batch effect was corrected using the algorithm ComBat (29) implemented in MSPrep.

### Statistical Analysis

Prior to analysis, metabolites data were log transformed and standardized to unit variance and zero mean ( $z$  scores). Continuous variables were also converted to standard normal distributions with corresponding mean and SD. Pearson partial correlation coefficients were calculated between identified metabolites and established clinical factors, adjusting for age, sex, and study site.

To identify metabolic predictors and to estimate their effects on the risk of developing T2D, we constructed a Cox proportional hazards frailty model, in which time to event was the dependent variable and the level of each metabolite was the independent variable. The frailty model was used here to account for the relatedness among family members. The proportional hazards assumption was tested using the Schoenfeld residuals, and it shows that the proportionality assumption holds in our data. For estimation of metabolic effects that are independent of traditional risk factors, the Cox frailty model was adjusted for age, sex, site, BMI, eGFR, HDL, triglycerides, fasting glucose, and insulin resistance (assessed by HOMA-IR) at baseline. Given the potential high correlations among detected metabolites, we used the  $q$  value method to adjust for multiple testing (30), and a  $q$  value <0.05 was considered statistically significant.

To examine the combined effects of metabolites on diabetes risk, we constructed a multimarker metabolites score based on metabolites that are significantly predictive of diabetes risk by fitting a model according to the following formula:  $\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$ , where  $X_i$  denotes the  $z$  score of the  $i$ -th metabolite and  $\beta_i$  denotes the regression coefficient from the logistic regression model containing the indicated metabolites. The joint predictive ability of metabolites was assessed using logistic regression by including all clinical risk factors (age, sex, study site, BMI, eGFR, HDL, triglycerides, fasting glucose, and HOMA-IR) plus the multimarker metabolite score compared with the model including clinical risk factors only. We calculated the area under the receiver operating characteristic curve (AUC), the net reclassification improvement (NRI), and the integrated discrimination improvement (IDI) to assess the incremental value of the metabolic markers for risk prediction beyond classical risk factors. Because our analysis was based on a regression model with no cross-validation or external validation, it is likely that our model could be overfitted. To avoid or minimize bias due to overfitting, we conducted a bootstrap estimation (1,000 reps) for coefficients by SAS to obtain bias-corrected estimates of metabolites on risk of diabetes.

To identify metabolic profiles associated with risk of diabetes, we conducted sparse partial least-squares discriminant analysis (sPLS-DA) using the computer package mixOmics implemented in R. The sPLS-DA is a supervised, multivariate technique to determine metabolic groups associated with disease risk. The sPLS-DA analysis included only metabolites showing significant associations with risk of diabetes. For ease of visualization, we presented a Manhattan plot ( $-\log_{10} P$  vs. metabolic feature) to show the significance of individual metabolites according to status of incident cases at follow-up using raw  $P$  values obtained from multivariate logistic regression analysis (false discovery rate at  $q = 0.05$  with a horizontal line).

### RESULTS

Table 1 presents the characteristics of the study participants at baseline (2001–2003) according to diabetes status at the end of follow-up (2006–2009). The average follow-up period was 5.5 years. Compared with participants who did not develop T2D, those who developed incident T2D had higher levels of BMI, triglycerides, fasting glucose, fasting insulin, and insulin resistance (HOMA-IR) but lower level of HDL at baseline. We also compared participants who were selected ( $n = 431$ ) versus those not selected ( $n = 1,686$ ) for this study. It shows that, except for BMI and eGFR, selected participants were not appreciably different from those not selected (Supplementary Table 1).

Our untargeted high-resolution LC-MS detected 11,628 distinct ions ( $m/z$ ) with CV  $\leq 10\%$ , of which 2,093  $m/z$  features matched known compounds in available metabolomics databases. Among all 11,628 features, altered levels of seven metabolites (five matching known metabolites and two unknown) were significantly associated with risk of diabetes after adjustment for clinical factors and multiple testing. Specially, a metabolite matching 2-hydroxybiphenyl (2HBP) and an unknown chemical ( $m/z$  ratio 1,178.804 [named X-1178]) were significantly associated with an increased risk of diabetes, whereas five metabolites matching phosphatidylcholine (PC 22:6/20:4), (3S)-7-hydroxy-2',3',4',5',8-pentamethoxyisoflavan (HPMF), two tetrapeptides (Met-Glu-Ile-Arg [MEIR] and

Leu-Asp-Tyr-Arg [LDYR]), and an unknown metabolite ( $m/z$  ratio 490.816 [named X-490]) were significantly associated with a decreased risk of diabetes. These associations are independent of clinical factors including fasting glucose and insulin resistance. Per-SD increase in the log-transformed levels of matching 2HBP and X-1178 was associated with 80% and 89%, respectively, increased risk of T2D. By contrast, per-SD increase in the log-transformed levels of matching PC (22:6/20:4), HPMF, tetrapeptides, and X-490 was associated with 32–42% decreased risk of T2D. In the multivariate model categorizing metabolites as tertiles, participants in the top tertile of 2HBP and X-1178 had a hazard ratio (HR) of 2.80 (95% CI 1.19–6.60) and 2.87 (95% CI 1.08–7.60) for developing incident T2D, respectively, compared with those in the lowest tertile. In contrast, participants in the top tertile of PC (22:6/20:4), HPMF, MEIR, LDYR, and X-490 had an HR of 0.45 (95% CI 0.21–0.97), 0.38 (95% CI 0.18–0.80), 0.44 (95% CI 0.20–0.96), 0.37 (95% CI 0.16–0.87), and 0.46 (95% CI 0.21–0.97) for developing T2D, respectively, compared with those in the lowest tertile of these metabolites.

To estimate the joint effects of metabolites on risk of diabetes development, we calculated HRs across tertiles of the combined metabolites comprising all seven significant metabolites. For the two risk metabolites (2HBP and X-1178), the HR for risk of developing incident T2D by comparing the top with the bottom tertiles of the summed metabolites was 6.89 (95% CI 2.63–18.08). For the five protective metabolites (PC [22:6/20:4], HPMF, MEIR, LDYR, and X-490), the HR of the top compared with the bottom tertiles of summed metabolites was 0.23 (95% CI 0.10–0.51). Multivariate associations of each individual metabolite along with their combined effects on diabetes risk are shown in Table 2. Of note, regression coefficients listed in Table 2 were corrected for potential overfitting by bootstrapping and thus should represent unbiased estimates of metabolic effects on risk of T2D. For ease of visual inspection, Fig. 1 shows a Manhattan plot ( $-\log_{10} P$  vs. metabolic feature) of all metabolites using raw  $P$  values obtained from multivariate regression analysis. Metabolites significantly predictive of diabetes risk are shown at the level of  $q = 0.05$ .

**Table 1—Characteristics of the study participants at baseline (2001–2003)**

	Participants who developed T2D	Participants who did not develop T2D	<i>P</i> *
<i>n</i>	133	298	
Age, years	35.45 ± 12.2	33.36 ± 13.88	0.1208
Female sex, %	67.67	63.42	0.3885
BMI, kg/m <sup>2</sup>	36.74 ± 7.96	31.11 ± 8.00	<0.0001
Current smoker, %	33.83	36.58	0.7266
Current drinker, %	63.16	68.79	0.5034
Systolic blood pressure, mmHg	120.88 ± 15.34	118.87 ± 12.96	0.1868
Diastolic blood pressure, mmHg	77.39 ± 11.80	75.63 ± 10.46	0.1222
HDL, mg/dL	47.52 ± 14.41	52.44 ± 14.63	0.0016
LDL, mg/dL	100.92 ± 29.32	96.06 ± 28.57	0.1062
Total triglyceride, mg/dL	167.20 ± 99.12	132.16 ± 65.47	<0.0001
Total cholesterol, mg/dL	180.70 ± 34.16	174.75 ± 33.48	0.0923
eGFR, mL/min/1.73 m <sup>2</sup>	104.56 ± 21.41	105.18 ± 24.84	0.7917
Fasting glucose, mg/dL	94.30 ± 7.81	89.55 ± 6.41	<0.0001
Fasting insulin, μU/mL	20.52 ± 13.08	14.14 ± 11.47	0.0001
Insulin resistance (HOMA-IR)	4.80 ± 3.07	3.15 ± 2.60	<0.0001
Total caloric intake, kcal/day	2,887.59 ± 2,079.25	2,812.91 ± 2,117.20	0.7409
Total dietary protein, g/day	97.51 ± 82.98	94.99 ± 81.77	0.7768
Total dietary fat, g/day	126.39 ± 99.66	123.71 ± 98.08	0.8017

Data are mean ± SD unless otherwise indicated. \*Adjusting for family relatedness by generalized estimating equation.

To investigate whether these detected metabolites improve risk prediction, we added the weighted multimarker score comprising all seven metabolites to the fully adjusted statistical model. Results show that addition of the metabolite score resulted in significant improvement for diabetes risk prediction as assessed by all three measures: the AUC value increased from 0.763 to 0.822 ( $P = 0.006$ ), the NRI was 0.623 (95% CI 0.427–0.819;  $P < 10^{-5}$ ), and the IDI was 0.117 (95% CI 0.083–0.151;  $P < 10^{-5}$ ). This indicates that the newly detected metabolic markers significantly improve

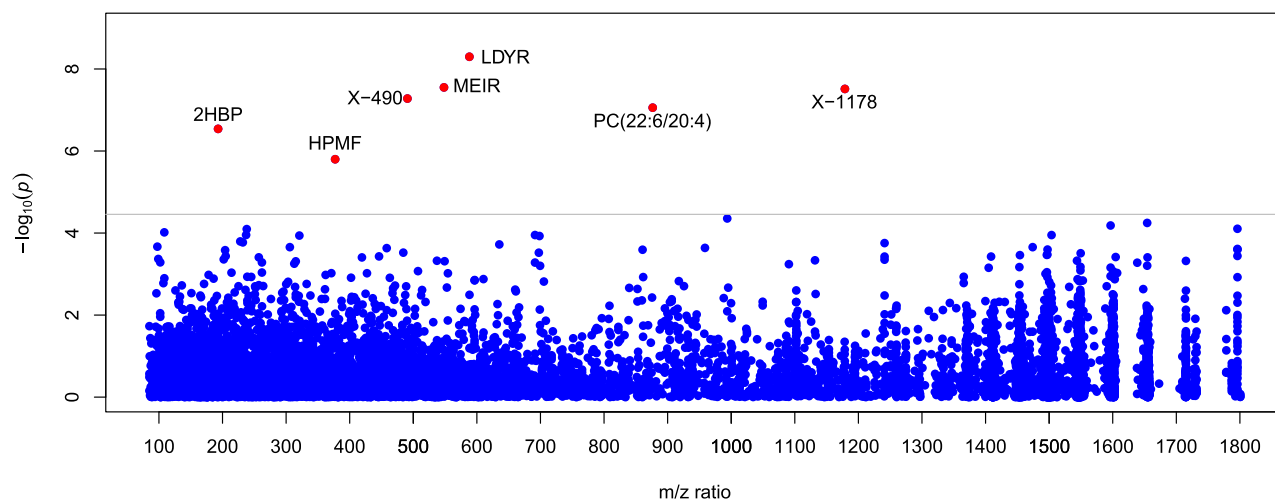
risk prediction of T2D beyond established diabetes risk factors. The five matching known metabolites belong to the classes of glycerophosphocholine, flavonoids, and polypeptides (Supplementary Table 2). Partial correlations of these matching metabolites with clinical risk factors are shown in Supplementary Table 3. Apart from some weak correlations of 2HBP with fasting insulin or insulin resistance, PC (22:6/20:4) with BMI, or LDYR with lipid levels, most metabolites were not correlated with established diabetes factors. The matching metabolites HPMF, MEIR, and the

**Table 2—Multivariate association of baseline fasting plasma metabolites with risk of developing T2D in American Indians by Cox proportional hazards frailty model†**

Matching metabolites	Metabolite as continuous variable*	Metabolite as categorical variable†*
Protective metabolites		
PC (22:6/20:4)	0.68 (0.52–0.88)	0.45 (0.21–0.97)
HPMF	0.58 (0.43–0.79)	0.38 (0.18–0.80)
MEIR	0.61 (0.47–0.78)	0.44 (0.20–0.96)
LDYR	0.63 (0.47–0.85)	0.37 (0.16–0.87)
X-490	0.65 (0.50–0.84)	0.46 (0.21–0.97)
Combined protective effects	0.43 (0.31–0.59)	0.23 (0.10–0.51)
Risk metabolites		
2HBP	1.80 (1.26–2.57)	2.80 (1.19–6.60)
X-1178	1.89 (1.29–2.77)	2.87 (1.08–7.60)
Combined risk effects	2.56 (1.71–3.84)	6.89 (2.63–18.08)

Data are HR (95% CI). †Adjusted for age, sex, site, BMI, eGFR, HDL, triglycerides, fasting glucose, and HOMA-IR. \*HR per SD change in log-transformed metabolite level. †Tertile 3 vs. tertile 1.





**Figure 1**—Manhattan plot of 11,628  $m/z$  features comparing participants who developed incident T2D versus those who did not. The negative log  $P$  value was plotted against the  $m/z$  features. The x-axis represents  $m/z$  of the detected features, ordered in increasing value from 85 (left) to 1,800 (right). A total of seven metabolites significantly differed between the two groups at the threshold of  $q = 0.05$  (above the horizontal gray line).

unknown compound (X-490) were not correlated with any of the known risk factors for diabetes.

To identify metabolic profiles associated with risk of diabetes development, we conducted sPLS-DA using the seven metabolites that were significantly predictive of disease risk. Fig. 2 demonstrates that participants who developed T2D and those who did not were separated into two distinct groups, suggesting that these metabolites could be used as discriminatory markers for T2D risk stratification. This observation is consistent with our results obtained by risk prediction analyses (i.e., AUC, NRI, and IDI). Additional adjustments for dietary intake of fat, protein, and caloric intake did not attenuate the observed associations (data not shown).

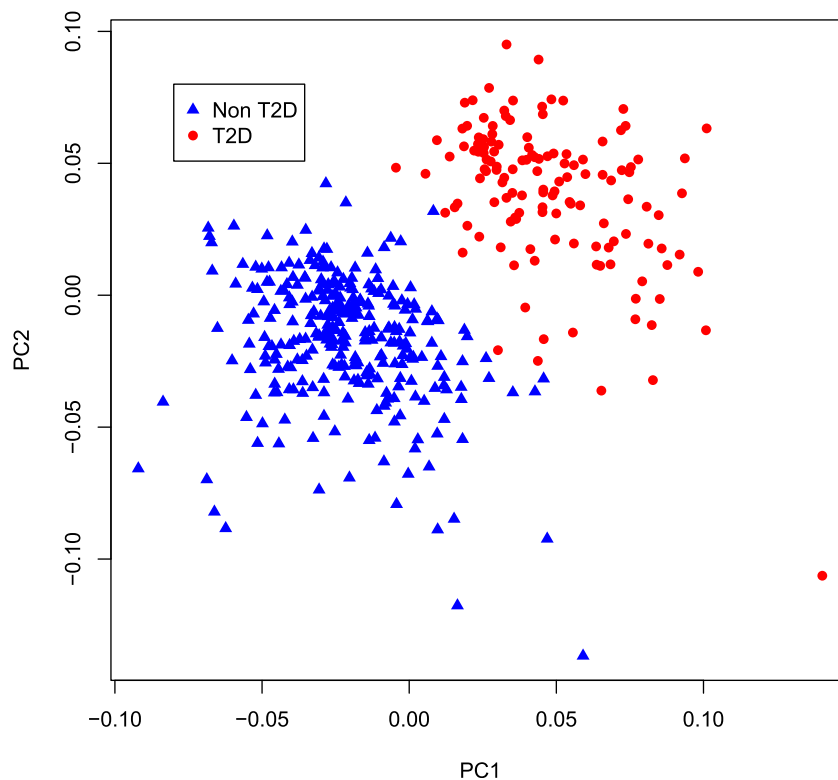
## CONCLUSIONS

In this prospective investigation using an untargeted high-resolution metabolomic approach, we found that seven metabolites independently predict future onset of T2D in American Indians, a group with a high rate of diabetes. Of the five chemicals matching known metabolites, two were lipids in the classes of glycerophosphocholine (PC) and flavonoid. It should be noted that there are many isobaric lipids, so the precise structural identifications will require additional research. The observed association withstood adjustments for multiple clinical indicators including age, sex, study site, BMI, eGFR, HDL, triglycerides, fasting glucose,

and insulin resistance (HOMA-IR). The combination of these metabolites significantly improves risk prediction beyond established diabetes risk factors. These metabolites have not been reported in previous studies of European individuals or other ethnic groups and thus should represent putative prognostic

markers of diabetes specific to American Indians.

We found that a metabolite matching 2HBP was associated with 80% increased risk of developing T2D independent of classical risk factors. The mechanism by which this metabolite affects diabetes risk is unclear. However,



**Figure 2**—Separation of study participants who developed incident T2D and those who did not during follow-up by sPLS-DA using a multimarker metabolite score comprising all seven metabolites showing significant associations with incident T2D listed in Table 2.

2HBP is known to be an environmental toxin that is widely used as industrial antimicrobials, agricultural fungicide, and disinfectants. 2HBP was reported to be mutagenic in human cells (31) and carcinogenic in animal models (32,33). In addition, hydroxybiphenyl chemicals can be degraded by bacteria through the biphenyl catabolic pathway (34). It is thus plausible to hypothesize that, apart from the possible direct toxic effects of 2HBP on pancreas or peripheral tissues, 2HBP may also negatively affect diabetes through a yet unknown host-gut microbiota mechanism.

Glycerophosphocholines are important structural components of plasma lipoproteins and cell membranes with diverse biological functions. In this study, we found that elevated plasma level of matching PC (22:6/20:4) was associated with 37% reduced risk of T2D in our study population. This is in agreement with a previous study demonstrating lower plasma or serum levels of PC species in diabetic patients than in control subjects (5). Moreover, reduced levels of multiple acyl-glycerophosphocholine species were highly correlated with insulin resistance as measured by the euglycemic clamp (35), lending further support for a potential role of PCs in diabetes etiology. In the current investigation, another metabolite matching known (3S)-7-hydroxy-2',3',4',5',8-pentamethoxyisoflavan (named HPMF) was also significantly predictive of a decreased risk of diabetes. This metabolite belongs to the class of flavonoids that are known to have a wide range of biological and pharmacological activities. Dietary flavonoid intakes have been associated with reduced risk of T2D in both human (36–38) and animal studies (39). In support of these findings, participants with a higher plasma level (top tertile) of HPMF exhibited over 60% reduced risk of T2D compared with those with a lower level (bottom tertile) in our analysis. While the precise mechanism underlying this association awaits further investigation, it is possible that HPMF may decrease diabetes risk through its potential antioxidant properties (40). It is also likely that HPMF may exert beneficial effects on energy balance and lipid metabolism (41) or anti-inflammatory effects through the nuclear factor- $\kappa$ B or the AMPK signaling pathways, which play a central role in the regulation of glucose

and lipid metabolism (42,43). In addition, flavonoids have been shown to have anti-diabetes effects through enhanced pancreatic  $\beta$ -cell function in animal experiments (44). The favorable effect of this flavonoid chemical has not been previously reported. Its biological properties should be investigated in future research.

In addition to the altered profiles of PC and flavonoid, elevated levels of two metabolites matching tetrapeptides (MEIR and LDYR) were associated with ~40% reduced risk of diabetes. Although the mechanisms linking these peptides to diabetes remain to be determined, peptides are known to be essential in regulating lipid metabolism in key insulin-target tissues and in maintaining energy homeostasis and insulin sensitivity. They may also function as potent peptide hormones regulating glucose metabolism in diabetes (45). In addition to the five known matching metabolites, two unknown compounds were also significantly predictive of diabetes development. These unknown chemicals might be not new but merely not yet identified. The structure and function of these unannotated chemicals should be examined in future research.

Previous evidence has linked raised circulating levels of BCAAs with insulin resistance (2,46,47) or diabetes (10,47). Our study, however, did not find a significant association of BCAAs with risk of T2D development. This lack of replication may not necessarily represent true negative findings because our analysis accounted for multiple testing of >11,000 *m/z* features with a stringent criterion, which could result in inappropriate exclusion for a large number of metabolites (false negatives). The discrepancy could also represent genuine difference between American Indians and other ethnic groups included in previous studies because the unique characteristics of American Indians, e.g., genetic background and lifestyle, could potentially lead to population-specific metabolic signatures. Future large-scale metabolomics studies should address this discrepancy.

In search of the origin of the interindividual variation, we calculated partial correlations of metabolite relative abundance with standard risk indicators of diabetes, expecting that, for example, higher BMI or fasting glucose should

correspond with higher levels of risk metabolites or lower levels of protective metabolites. However, in our study cohort, most of the detected matching metabolites were not correlated with classical risk factors, such as BMI, fasting glucose, and insulin resistance, but the combination of these metabolites significantly improved risk prediction beyond standard risk factors. This is important because the fundamental task of risk prediction is to identify predictive markers that are sufficiently uncorrelated with established risk factors so that they can be used to improve risk prediction over and above conventional clinical factors. These newly detected metabolic markers will provide valuable information regarding the pathophysiology of diabetes development and also potential therapeutic targets for novel treatment options.

Our study has several limitations. First, although our high-resolution LC-MS detected >11,000 distinct features, it should be noted that only 18% of the compounds detected had a match in the current metabolomics database. These compounds were unable to be pursued owing to the large number of possible isomers and a lack of available standards. However, these currently unannotated metabolites may represent dietary, microbiome-related, or environmental chemicals associated with diabetes. With the advancement of metabolomic research, we expect that the majority of these unidentified chemicals will ultimately be annotated and their associations with disease will be determined. Additionally, many *m/z* features matched to therapeutic drugs and nutritional supplements, but owing to their wide use by diabetic patients, we were unable to evaluate their contributions to the altered metabolic profiles. Second, although highly correlated, relative abundances but not absolute concentrations were used as a surrogate for plasma metabolite levels. Third, although we were able to control many of the known risk factors, the possibility of potential confounding by other factors such as diet and gut microbiota cannot be entirely excluded. Fourth, participants in the current study are young to middle-aged American Indians who may have a high propensity for the development of T2D; therefore, generalization of our findings to other populations should be approached cautiously. However, given the rising tide of T2D in

almost all ethnic groups worldwide, we believe that our results could be applicable to other populations. Finally, our results need to be replicated in large-scale, prospective metabolomic analysis of American Indians and other ethnic groups.

Nonetheless, this is the first prospective study to report novel predictive metabolic markers and altered metabolic profiles associated with development of T2D in American Indians, a minority group suffering from a disproportionately high rate of T2D. The SHFS has phenotypic longitudinal data available that allowed us to accurately classify participants as incident cases of diabetes. The untargeted high-resolution metabolomics approach allowed us to identify previously undescribed metabolic markers that may be specific to the population of American Indians, whose genetic makeup and/or lifestyle could be distinct from that of individuals of European ancestry.

In summary, this study identified significant metabolic predictors of T2D in American Indians above and over established diabetes indicators. Targeting biological pathways that involve these newly detected metabolites would help to develop early preventive and therapeutic strategies tailored to American Indians, an ethnically important but traditionally understudied minority population.

**Acknowledgments.** The authors thank the SHFS participants, Indian Health Service facilities, and participating tribal communities for their extraordinary cooperation and involvement, which has contributed to the success of the SHFS.

**Funding.** This study was supported by National Institutes of Health grants R01DK091369, K01AG034259, and R21HL092363 and cooperative agreement grants U01HL65520, U01HL41642, U01HL41652, U01HL41654, and U01HL65521.

The views expressed in this article are those of the authors and do not necessarily reflect those of the Indian Health Service.

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

**Author Contributions.** J.Z. conceived the study, supervised the statistical analyses, and wrote the manuscript. Y.Z., N.H., and D.Z. conducted statistical analyses. K.U. and V.T.T. collected LC-MS data and conducted metabolomic analyses. T.Y. and D.J. supervised metabolomic data analyses. J.H., E.T.L., and B.V.H. contributed to study design, data interpretation, and discussion and reviewed and edited the manuscript. J.Z. is the guarantor of this work and, as such, had full access to all of the data in the study and takes responsibility for

the integrity of the data and the accuracy of the data analysis.

## References

1. DeFronzo RA. Banting Lecture. From the trimvirate to the ominous octet: a new paradigm for the treatment of type 2 diabetes mellitus. *Diabetes* 2009;58:773–795
2. Newgard CB, An J, Bain JR, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab* 2009;9:311–326
3. Würtz P, Mäkinen VP, Soininen P, et al. Metabolic signatures of insulin resistance in 7,098 young adults. *Diabetes* 2012;61:1372–1380
4. Wopereis S, Rubingh CM, van Erk MJ, et al. Metabolic profiling of the response to an oral glucose tolerance test detects subtle metabolic changes. *PLoS ONE* 2009;4:e4525
5. Suhre K, Meisinger C, Döring A, et al. Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. *PLoS ONE* 2010;5:e13953
6. Adams SH, Hoppel CL, Lok KH, et al. Plasma acylcarnitine profiles suggest incomplete long-chain fatty acid beta-oxidation and altered tricarboxylic acid cycle activity in type 2 diabetic African-American women. *J Nutr* 2009;139:1073–1081
7. Fiehn O, Garvey WT, Newman JW, Lok KH, Hoppel CL, Adams SH. Plasma metabolomic profiles reflective of glucose homeostasis in non-diabetic and type 2 diabetic obese African-American women. *PLoS ONE* 2010;5:e15234
8. Wang-Sattler R, Yu Z, Herder C, et al. Novel biomarkers for pre-diabetes identified by metabolomics. *Mol Syst Biol* 2012;8:615
9. Menni C, Fauman E, Erte I, et al. Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach. *Diabetes* 2013;62:4270–4276
10. Wang TJ, Larson MG, Vasani RS, et al. Metabolite profiles and the risk of developing diabetes. *Nat Med* 2011;17:448–453
11. 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
12. 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
13. North KE, Howard BV, Welty TK, et al. Genetic and environmental contributions to cardiovascular disease risk in American Indians: the strong heart family study. *Am J Epidemiol* 2003;157:303–314
14. Lee ET, Welty TK, Fabsitz R, et al. The Strong Heart Study. A study of cardiovascular disease in American Indians: design and methods. *Am J Epidemiol* 1990;132:1141–1155
15. Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* 2003;26(Suppl. 1):S5–S20
16. Clauss A. Rapid physiological coagulation method in determination of fibrinogen. *Acta Haematol* 1957;17:237–246 [in German]
17. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 1985;28:412–419
18. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D; Modification of Diet in Renal Disease Study Group. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. *Ann Intern Med* 1999;130:461–470
19. Willett W. Food frequency methods. In *Nutritional Epidemiology*. 2nd ed. New York, Oxford University Press, 1998, p. 74–91
20. Osborn MP, Park Y, Parks MB, et al. Metabolome-wide association study of neovascular age-related macular degeneration. *PLoS ONE* 2013;8:e72737
21. Roede JR, Uppal K, Park Y, et al. Serum metabolomics of slow vs. rapid motor progression Parkinson's disease: a pilot study. *PLoS ONE* 2013;8:e77629
22. Yu T, Park Y, Johnson JM, Jones DP. apLCMS—adaptive processing of high-resolution LC/MS data. *Bioinformatics* 2009;25:1930–1936
23. Yu T, Park Y, Li S, Jones DP. Hybrid feature detection and information accumulation using high-resolution LC-MS metabolomics data. *J Proteome Res* 2013;12:1419–1427
24. Uppal K, Soltow QA, Strobel FH, et al. xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. *BMC Bioinformatics* 2013;14:15
25. Smith CA, O'Maille G, Want EJ, et al. METLIN: a metabolite mass spectral database. *Ther Drug Monit* 2005;27:747–751
26. Wishart DS, Jewison T, Guo AC, et al. HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Res* 2013;41:D801–D807
27. Sud M, Fahy E, Cotter D, et al. LMSD: LIPID MAPS structure database. *Nucleic Acids Res* 2007;35:D527–D532
28. Hughes G, Cruickshank-Quinn C, Reisdorph R, et al. MSPrep—summarization, normalization and diagnostics for processing of mass spectrometry-based metabolomic data. *Bioinformatics* 2014;30:133–134
29. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007;8:118–127
30. Storey JD. A direct approach to false discovery rates. *J R Stat Soc Series B Stat Methodol* 2002;64:187–205
31. Suzuki H, Suzuki N, Sasaki M, Hiraga K. Orthophenylphenol mutagenicity in a human cell strain. *Mutat Res* 1985;156:123–127
32. Brusick D. Analysis of genotoxicity and the carcinogenic mode of action for ortho-phenylphenol. *Environ Mol Mutagen* 2005;45:460–481
33. Hagiwara A, Shibata M, Hirose M, Fukushima S, Ito N. Long-term toxicity and carcinogenicity study of sodium o-phenylphenate in B6C3F1 mice. *Food Chem Toxicol* 1984;22:809–814
34. Sondossi M, Sylvestre M, Ahmad D, Masse R. Metabolism of hydroxybiphenyl and chlorohydroxybiphenyl by biphenyl/chlorobiphenyl degradign *Pseudomonas testosteroni*, strain B-356. *J Ind Microbiol* 1991;6:77–88

35. Gall WE, Beebe K, Lawton KA, et al.; RISC Study Group. alpha-Hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a nondiabetic population. *PLoS ONE* 2010;5:e10883
36. van Dam RM, Naidoo N, Landberg R. Dietary flavonoids and the development of type 2 diabetes and cardiovascular diseases: review of recent findings. *Curr Opin Lipidol* 2013;24:25–33
37. Wedick NM, Pan A, Cassidy A, et al. Dietary flavonoid intakes and risk of type 2 diabetes in US men and women. *Am J Clin Nutr* 2012;95:925–933
38. Zamora-Ros R, Forouhi NG, Sharp SJ, et al. The association between dietary flavonoid and lignan intakes and incident type 2 diabetes in European populations: the EPIC-InterAct study. *Diabetes Care* 2013;36:3961–3970
39. Chen YK, Cheung C, Reuhl KR, et al. Effects of green tea polyphenol (-)-epigallocatechin-3-gallate on newly developed high-fat/Western-style diet-induced obesity and metabolic syndrome in mice. *J Agric Food Chem* 2011;59:11862–11871
40. Lotito SB, Zhang WJ, Yang CS, Crozier A, Frei B. Metabolic conversion of dietary flavonoids alters their anti-inflammatory and antioxidant properties. *Free Radic Biol Med* 2011;51:454–463
41. Friedrich M, Petzke KJ, Raederstorff D, Wolfram S, Klaus S. Acute effects of epigallocatechin gallate from green tea on oxidation and tissue incorporation of dietary lipids in mice fed a high-fat diet. *Int J Obes (Lond)* 2012;36:735–743
42. Salminen A, Hyttinen JM, Kaarniranta K. AMP-activated protein kinase inhibits NF- $\kappa$ B signaling and inflammation: impact on healthspan and lifespan. *J Mol Med (Berl)* 2011;89:667–676
43. Leiherer A, Mündlein A, Drexel H. Phytochemicals and their impact on adipose tissue inflammation and diabetes. *Vascul Pharmacol* 2013;58:3–20
44. Ortsäter H, Grankvist N, Wolfram S, Kuehn N, Sjöholm A. Diet supplementation with green tea extract epigallocatechin gallate prevents progression to glucose intolerance in db/db mice. *Nutr Metab (Lond)* 2012;9:11
45. Todd JF, Bloom SR. Incretins and other peptides in the treatment of diabetes. *Diabet Med* 2007;24:223–232
46. Huffman KM, Shah SH, Stevens RD, et al. Relationships between circulating metabolic intermediates and insulin action in overweight to obese, inactive men and women. *Diabetes Care* 2009;32:1678–1683
47. 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