Geoffrey A. Walford,[1,2,3] Yong Ma,[4,5] Clary Clish,[6] Jose C. Florez,[1,2,3] Thomas J. Wang,[7] and Robert E. Gerszten[3,6,8,9] for the Diabetes Prevention Program Research Group*

Metabolite Profiles of Diabetes Incidence and Intervention Response in the Diabetes Prevention Program

Identifying novel biomarkers of type 2 diabetes risk may improve prediction and prevention among individuals at high risk of the disease and elucidate new biological pathways relevant to diabetes development. We performed plasma metabolite profiling in the Diabetes Prevention Program (DPP), a completed trial that randomized high-risk individuals to lifestyle, metformin, or placebo interventions. Previously reported markers, branched-chain and aromatic amino acids and glutamine/glutamate, were associated with incident diabetes ($P < 0.05$ for all), but these associations were attenuated upon adjustment for clinical and biochemical measures. By contrast, baseline levels of betaine, also known as glycine betaine (hazard ratio 0.84 per SD log metabolite level, $P = 0.02$), and three other metabolites were associated with incident diabetes even after adjustment. Moreover, betaine was increased by the lifestyle intervention, which was the most effective approach to preventing diabetes, and increases in betaine at 2 years were also associated with lower diabetes incidence ($P = 0.01$). Our findings indicate betaine is a marker of diabetes risk among high-risk individuals both at baseline and during preventive interventions and they complement animal models demonstrating a direct role for betaine in modulating metabolic health.

Early interventions aimed at the prevention of type 2 diabetes can be effective when targeted at individuals at high risk (1). Elucidating novel risk markers may highlight new biology relevant to the transition from a high-risk metabolic state to overt type 2 diabetes and also permit the early identification of high-risk individuals who are likely to benefit or fail preventive interventions for type 2 diabetes.

Higher circulating concentrations of branched-chain amino acids and aromatic amino acids (BCAA/As) (isoleucine, leucine, valine, tyrosine, and phenylalanine) (2,3) and a lower ratio of circulating glutamine-to-glutamate concentrations (glutamine/glutamate) (4,5) are observed in individuals who develop type 2 diabetes up to 12 years into the future. However, prior studies have been conducted in observational cohorts of individuals, who were primarily of European background and at average risk for type 2 diabetes.

The Diabetes Prevention Program (DPP) was a multiethnic, randomized clinical trial of lifestyle or pharmacological interventions to prevent diabetes in high-risk overweight or obese individuals (6). All participants had elevated fasting glucose and impaired glucose tolerance. The main results of the DPP were that an intensive lifestyle intervention and metformin reduced diabetes incidence relative to placebo (1). Using stored samples from the DPP, we performed metabolite profiling to test the hypotheses that previously established amino acid markers (BCAA/As and glutamine/glutamate) can predict diabetes incidence in a high-risk population and can inform the

GENETICS/GENOMES/PROTEOMICS/METABOLOMICS

[1]Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA
[2]Diabetes Clinical Research Center (Diabetes Unit), Department of Medicine, Massachusetts General Hospital, Boston, MA
[3]Harvard Medical School, Boston, MA
[4]The George Washington University Biostatistics Center, Rockville, MD
[5]Department of Epidemiology and Biostatistics, Milken Institute School of Public Health, The George Washington University, Washington, DC
[6]Metabolomics Platform, Broad Institute, Cambridge, MA
[7]Division of Cardiology, Vanderbilt University Medical Center, Nashville, TN
[8]Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA
[9]Cardiology Division, Massachusetts General Hospital, Boston, MA

Corresponding author: Geoffrey A. Walford, dppmail@bsc.gwu.edu.

response to preventive interventions for type 2 diabetes. We also expanded our analyses to assess the relationship between additional circulating metabolites and diabetes incidence.

RESEARCH DESIGN AND METHODS

DPP

The DPP Research Group conducted a randomized, multi-ethnic clinical trial from 1996 to 2001 at 27 centers in the U.S. (1). Full inclusion and exclusion criteria are detailed in the published DPP study design (6). Briefly, all participants had elevated BMI, elevated fasting glucose, and impaired glucose metabolism but not diabetes at baseline.

During the main phase of the trial, 3,234 participants were followed for an average of 3.2 years after random assignment to one of three intervention groups: 1,079 to intensive lifestyle, 1,073 to metformin, and 1,082 to placebo (1). Participants in the intensive lifestyle group engaged in a healthy, low-calorie, low-fat diet and 150 min of brisk exercise per week to achieve and maintain a goal of at least 7% weight reduction from baseline. The main findings of the DPP have been published: compared with placebo, the intensive lifestyle and metformin interventions reduced the incidence of diabetes by 58% and 31%, respectively, after a mean 2.8 years of follow-up (1).

The institutional review board at each center approved the protocol, and all participants gave written informed consent.

Sample Selection

Sample selection for metabolite profiling followed a nested case-control design with 1:1 matching of participants with and without incident diabetes during the DPP with data lock at 31 July 2001. All racial/ethnic groups were included, and metabolite samples were drawn from all of the DPP study centers. For each individual with incident diabetes, at the time of the diabetes assessment, an individual free of diabetes was selected by matching sex, treatment group (lifestyle, metformin, or placebo), hypertension status, and propensity score. The propensity score was calculated at each 6-month assessment for diabetes within the DPP (1 to 9, indicating 6 months to 54 months) from a logistic regression model with the diabetes status (yes/no) as the outcome and age at randomization, race/ethnicity, baseline fasting plasma glucose (FPG), and BMI as predictors. The propensity score was calculated by the following equation: logit (diabetes status) = age at randomization + race/ethnicity + baseline FPG + BMI. Ten strata were created from the propensity score by rounding the score (0 to 1) to the nearest 0.1. A participant free of diabetes may have been used as a matched control subject at multiple time points during the study. Individuals serving as matched control subjects earlier could have developed diabetes at a later time. In total, there were 427 pairings of individuals with and without incident diabetes; however, because some individuals without incident diabetes were used in multiple pairings only 330 individuals without incident diabetes actually contributed to the pairings. Fasting samples from these 757 DPP participants were used for metabolite measurements.

Metabolite Measurements

Measurements were performed on stored samples at baseline and 2 years of follow-up in the placebo, intensive lifestyle, and metformin groups. Previously aliquoted plasma samples stored at −80°C since collection during the DPP trial were thawed and centrifuged for 5 min at 1,500 rpm. All samples were exposed to one freeze-thaw cycle before measurement.

Metabolite profiling was performed as previously described (2,7,8). Briefly, we used hydrophilic interaction liquid chromatography coupled to tandem mass spectrometry (MS) to analyze polar metabolites in the positive ion mode in a targeted manner. This platform assesses amino acids, amines, acylcarnitines, some nucleotides, and other compound classes. All liquid chromatography–MS analyses were performed using a 4000 QTRAP triple quadrupole mass spectrometer (AB Sciex, Foster City, CA) coupled to either an 1100 Series pump (Agilent Technologies, Santa Clara, CA) or an HTS PAL autosampler (Leap Technologies, Carrboro, NC) equipped with a column heater.

Quality control measures were implemented. Prior to each run, mixtures of reference standards containing 130 metabolites were measured to assure reproducibility of chromatographic retention times, peak shapes, and MS sensitivity. During each run, isotope-labeled internal standards were monitored in each sample to ensure proper injection and to monitor MS sensitivity. Further, a pooled plasma reference sample was inserted in the analysis queue after every 20 study samples. Thus, the pooled plasma sample was used as a reference to standardize within and across batches by "nearest neighbor" scaling. The coefficient of variance for each metabolite was calculated to determine the reproducibility over the run, with close monitoring for temporal drift. We have previously demonstrated that median coefficient of variances are ≤20% for 74% of metabolites and ≤10% for 54% of metabolites, including BCAA/As (2). No study site or batch effects were noted. We have successfully used this "double standardization" technique (to isotope in the same class of the analyte of interest and to the corresponding metabolite in the "nearest neighbor" pooled plasma) to control for batch effects in our prior biomarker studies in multiple cohorts (2,4,9). Finally, each batch contained random samples from the three treatment groups. All metabolite measurements were normalized to pooled plasma samples and internal standards and then log transformed.

In total, 79 of 84 metabolites were measured in >75% of the study cohort and were included in the analyses. The complete list of 84 metabolites measured in this study is provided in Supplementary Table 1. Glutamine/glutamate, which has been associated with type 2 diabetes (4,5), was also analyzed as a prespecified metabolite.

Anthropomorphic and Clinical Laboratory Measurements

Weight and height were used to calculate BMI. Waist circumference was measured in centimeters. Laboratory measurements for FPG and fasting insulin were performed by glucokinase and radioimmunoassay methods, respectively, during the DPP study (6,10). Diabetes onset was detected by an oral glucose tolerance test at an annual visit or by a semiannual fasting glucose level, with confirmation on a second test within 6 weeks (6), using the 1999 American Diabetes Association criteria (11).

Statistical Analysis

All metabolite values were log transformed, which led to an approximately normal distribution for each metabolite frequency and reduced heteroscedasticity of selected samples. Percent metabolite change was calculated using the log-transformed values at 2 years of intervention and baseline.

Due to prior sample use for other biomarker studies in the DPP, the percentage of stored samples missing for the metabolite measurement varied across the treatment groups and by incident diabetes status (35% missing samples for individuals who developed diabetes vs. 22% missing samples for those who remained free of diabetes; 36% missing samples in the placebo group vs. 22% and 17% missing samples in the lifestyle and metformin groups, respectively). Thus, to correct for potential biases introduced by the nonrandom missingness of samples, we applied an inverse probability weighting (IPW) approach to analyze the case-control study using established methods that facilitate unmatched analysis of matched data (12,13). Logistic regression models were built separately for selected individuals with incident diabetes and selected individuals without incident diabetes to create a propensity score for participants who were selected for metabolite profiling. The inverses of the propensity scores were then used as the weights in the subsequent analyses. The primary study results of treatment effect in the DPP as well as baseline characteristics of the DPP cohort were restored after using the IPW (Table 1), demonstrating the effectiveness of the method.

Weighted Cox proportional hazard models were used to examine the association between metabolites and risk of diabetes during the DPP (up to 31 July 2001). Variances of the parameter estimates were calculated using the Jackknife resampling method. Interaction between treatment arm and baseline metabolite or metabolite change was tested for each metabolite. If the interaction term was significant ($P < 0.05$), the analyses were stratified and presented by treatment arm. If the interaction term was not significant, the interaction term was removed and the model was reanalyzed with only a parameter for treatment effect.

Models were analyzed without and with adjustment for treatment group, sex, race/ethnicity, and baseline age, BMI, FPG, and hypertension status to assess the contribution of metabolite values to standard clinical predictors of diabetes. Models examining the relationship between metabolite change and diabetes incidence were also adjusted for baseline metabolite level and included only participants free of diabetes at the 2-year follow-up assessment. We also performed conditional logistic regression with the matched case-control data, and the results were similar to the IPW approach.

All analyses were conducted using SAS (version 9.2; Cary, NC). Two-sided *P* values are reported. Statistical significance was set at $\alpha = 0.05$ for the primary analyses examining the association of the established amino acid markers, BCAA/As and glutamine/glutamate, with incident diabetes. We also report the results for additional metabolites that reach this statistical threshold in exploratory analyses.

RESULTS

Table 1 shows the baseline characteristics of the DPP cohort, the individuals with and without incident diabetes who were selected for metabolite profiling, and the metabolite profiling cohort of the DPP after the application of IPW, an established statistical method to correct for potential biases introduced by sample availability (12,13).

Owing to the DPP inclusion criteria, all participants had impaired glucose tolerance based on an oral glucose tolerance test, elevated BMI, and elevated fasting glucose. Relative to the original DPP cohort, the selected cohort of individuals with and without incident diabetes had more participants in the metformin group and fewer participants in the lifestyle intervention group, were slightly older, and had fewer men and more women. After the application of IPW, the characteristics of the metabolite profiling cohort replicated those in the original DPP cohort, and baseline characteristics across interventional treatment groups were balanced. Further, the relative reduction in diabetes incidence in the lifestyle (56%) and metformin (35%) arms relative to placebo in the IPW data set closely approximated that in the DPP full cohort (1). Therefore, to present findings that represented those in the full DPP cohort, all reported results use the IPW approach unless otherwise stated.

Incident diabetes occurred at an average of 23 months (range 6–54 months) after randomization in the DPP. Among those individuals free of diabetes at the 2-year metabolite assessment, incident diabetes occurred at an average of 13 months (range 0–30 months) later during the DPP.

Metabolite profiling was performed according to previously described methods (2,7,8). All measured metabolite values (Supplementary Table 1) were log transformed, which led to an approximately normal distribution for each metabolite frequency and reduced heteroscedasticity of selected samples.

Table 2 demonstrates the relationship of baseline levels of the established amino acid markers (BCAA/As and glutamine/glutamate) with incident diabetes in

Table 1—Baseline characteristics of the DPP, individuals with and without incident diabetes who were selected for metabolite profiling, and metabolite profiling cohort after application of IPW

	DPP			IPW			
	Entire DPP cohort	Selected with incident diabetes	Selected without incident diabetes	Entire DPP cohort	Lifestyle	Metformin	Placebo
N (entire cohort)	3,234	427	330*	3,180	993	1,097	1,090
Lifestyle group	1,079 (33.4)	99 (23.2)	99 (23.2)	993 (31.2)			
Metformin group	1,073 (33.2)	185 (43.3)	185 (43.3)	1,097 (34.5)			
Placebo group	1,082 (33.4)	143 (33.5)	143 (33.5)	1,090 (34.3)			
Age, years	50.6 (10.70)	51.89 (9.84)	51.36 (10.06)	50.7 (0.51)	51.4 (1.01)	51.1 (0.88)	49.7 (0.74)
Sex							
Male	1,043 (32.3)	112 (26.2)	112 (26.2)	979 (30.8)	349 (35.2)	457 (41.7)	173 (15.9)
Female	2,191 (67.7)	315 (73.8)	315 (73.8)	2,201 (69.2)	644 (64.8)	640 (58.3)	917 (84.1)
Race/ethnicity							
Caucasian	1,768 (54.7)	233 (54.6)	220 (51.5)	1,757 (55.3)	499 (50.2)	658 (60.0)	601 (55.1)
African American	645 (19.9)	82 (19.2)	97 (22.7)	649 (20.4)	239 (24.1)	157 (14.3)	252 (23.2)
Hispanic	508 (15.7)	61 (14.3)	55 (12.9)	479 (15.1)	149 (15.0)	179 (16.3)	151 (13.8)
Asian	142 (4.4)	21 (4.9)	26 (6.1)	140 (4.4)	58 (5.8)	44 (4.0)	38 (3.5)
American Indian	171 (5.3)	30 (7.0)	29 (6.8)	155 (4.9)	48 (4.8)	58 (5.3)	49 (4.5)
BMI, kg/m²	34.0 (6.70)	34.41 (6.70)	34.74 (6.72)	34.5 (0.38)	35.4 (0.73)	33.5 (0.47)	34.7 (0.73)
Waist circumference, cm	105.1 (14.50)	106.5 (15.62)	106.0 (14.21)	105.5 (0.76)	107.8 (1.77)	104.8 (0.96)	104.3 (1.14)
Smoking	226 (7.0)	35 (8.2)	21 (4.9)	170 (5.3)	47 (4.7)	61 (5.6)	61 (5.6)
Hypertensive							
No	2,309 (71.4)	285 (66.7)	285 (66.7)	2,254 (70.9)	729 (73.5)	719 (65.5)	805 (73.9)
Yes	925 (28.6)	142 (33.3)	142 (33.3)	926 (29.1)	263 (26.5)	378 (34.5)	285 (26.1)
High LDL medication use							
No	3,074 (95.1)	407 (95.3)	399 (93.4)	2,988 (93.9)	945 (95.1)	1,019 (92.9)	1,025 (94.0)
Yes	160 (4.9)	20 (4.7)	28 (6.6)	192 (6.1)	48 (4.9)	78 (7.1)	66 (6.0)
Statin medication use¥	137 (4.2)	17 (4.0)	27 (6.4)	183 (5.8)	47 (4.7)	73 (6.7)	62 (5.7)
Nonstatin medication use¥	23 (0.7)	3 (0.7)	1 (0.2)	9 (0.3)	1 (0.2)	5 (0.2)	4 (0.3)
Family history of diabetes							
No	991 (30.6)	122 (28.6)	121 (28.5)	959 (30.2)	313 (31.5)	320 (29.1)	327 (30.0)
Yes	2,243 (69.4)	304 (71.4)	304 (71.5)	2,219 (69.8)	680 (68.5)	777 (70.9)	762 (70.0)
Leisure activity, MET-h¶	16.3 (25.8)	16.90 (26.1)	15.73 (20.3)	16.3 (1.1)	15.4 (2.4)	18.3 (2.0)	15.1 (1.3)
Daily caloric intake, kcal	2,126 (1,037.0)	2,104 (1,001.0)	2,134 (1,012.0)	2,075 (46.6)	2,065 (95.1)	2,152 (70.7)	2,009 (75.6)
Daily protein intake, g	88.7 (44.30)	88.70 (41.27)	89.88 (43.96)	86.8 (1.98)	84.1 (3.82)	90.3 (3.17)	85.7 (3.27)
FPG, mg/dL	106.5 (8.30)	110.0 (8.90)	109.5 (9.06)	106.6 (0.42)	107.6 (0.91)	106.7 (0.54)	105.8 (0.72)
2-h Plasma glucose, mg/dL‡	164 (17.0)	171.9 (17.4)	163.9 (17.3)	163.8 (0.9)	163.2 (1.7)	164.9 (1.3)	163.2 (1.5)

Table 1—Continued

	DPP			IPW			
	Entire DPP cohort	Selected with incident diabetes	Selected without incident diabetes	Entire DPP cohort	Lifestyle	Metformin	Placebo
Serum triglycerides, mg/dL	163.4 (95.9)	186.0 (110.1)	161.1 (93.3)	159.1 (3.9)	151.2 (8.7)	162.6 (5.1)	162.7 (6.3)
Fasting insulin, mU/L	26.7 (15.2)	29.18 (16.1)	27.01 (13.0)	26.5 (0.7)	26.7 (1.5)	26.9 (0.9)	25.8 (1.0)
HbA_{1c}, % [mmol/mol]	5.91 (0.50) [41 (4.3)]	6.1 (0.50) [43 (4.3)]	6.0 (0.50) [42 (4.3)]	5.91 (0.30) [41 (2.6)]	5.92 (0.60) [41 (5.2)]	5.88 (0.30) [41 (2.6)]	5.93 (0.40) [41 (2.96)]

Data shown are mean (SD) for all quantitative variables and n (%) for categorical variables, unless stated otherwise. The DPP shows baseline characteristics of the DPP cohort. Individuals with and without incident diabetes selected for metabolite profiling were matched on treatment group allocation, sex, and hypertension status, as well as propensity score calculated using age, race/ethnicity, BMI, and FPG. *In total, there were 427 pairings of individuals with and without incident diabetes; however, because some individuals without incident diabetes were used in multiple pairings only 330 individuals without incident diabetes actually contributed to the pairings. Following implementation of the IPW, the demographic characteristics of the DPP were restored in the IPW DPP cohort, which was used for all analyses. Baseline characteristics were balanced across all treatment groups after IPW. More females than males participated in the DPP and this imbalance is demonstrated in the IPW cohort as well. Hypertensive status included participants with known diagnosis and those on blood pressure–lowering medication. ¥Number and percentage of subjects using statin medications and using nonstatin medications are provided based on total number of subjects. ¶MET-h represents the average amount of time engaged in specific physical activities multiplied by the MET (metabolic equivalent) value of each activity and is based on responses to the Modifiable Activity Questionnaire. ‡2-h Plasma glucose determined after a 75-g oral glucose load.

the DPP. Prior to adjustment for baseline characteristics, each established amino acid marker except phenylalanine was associated with increased diabetes incidence during the DPP ($P < 0.05$). After adjustment for age, sex, race/ethnicity, treatment group, BMI, FPG, and hypertension status at baseline, the association of diabetes incidence with each established amino acid marker, except phenylalanine, remained increased as expected, but was no longer significant. The inclusion of FPG measurement in the models led to the largest attenuation of the association between the amino acids and incident diabetes. Adjusting the models for other covariates, including treatment group, race/ethnicity, age, sex, BMI, and hypertension status, did not markedly attenuate the relationship between amino acids and incident diabetes. Additional adjustment for baseline smoking status in these models also did not change the associations with incident diabetes (data not shown).

We next performed analyses across all metabolites measured by the platform to identify additional metabolites markers of diabetes incidence. In analyses that accounted for baseline characteristics, we found four metabolites associated with incident diabetes in the DPP: methionine sulfoxide, betaine (also known as glycine betaine), serine, and propionylcarnitine (Table 3). Additional adjustment for baseline smoking status in these models also did not change the associations with incident diabetes (data not shown). Serine, betaine, and methionine sulfoxide are all metabolites within the choline metabolism pathway. Baseline levels of betaine and serine were correlated with each other ($P < 0.0001$), with methionine sulfoxide ($P < 0.005$ for both), and with propionylcarnitine ($P < 0.0005$ for both), yet baseline levels of these metabolites were not correlated with baseline BMI, FPG, or fasting insulin (Fig. 1). Further, repeating the statistical models without adjustment for FPG did not change the hazard ratios (HRs) (95% CI) for the relationship between these metabolites and incident diabetes (methionine sulfoxide [1.11 (1.01, 1.24)]; betaine [0.86 (0.74, 1.02)]; serine [0.91 (0.78, 1.05)]; and propionylcarnitine [1.23 (1.01, 1.55)]), supporting that the relationship of these metabolites with incident diabetes was not dependent on measured glucose levels.

To assure bias was not introduced by the IPW approach, we performed conditional logistic regression models using only the matched nested case-control cohort (i.e., using only directly measured samples and no IPW). Because these models used participants matched on baseline characteristics, we expected the results to be similar to those of the IPW models adjusted for age, sex, race/ethnicity, treatment group, BMI, FPG, and hypertension status. Indeed, no significant associations were observed in the matched case-control cohort between incident diabetes and established amino acid risk markers (HR [95% CI] per SD increase in log metabolite level for isoleucine 1.23 [0.72, 2.10], $P = 0.45$; leucine 1.30 [0.69, 2.46], $P = 0.42$; valine 1.25 [0.63, 2.48], $P = 0.53$;

Table 2—Established amino acids markers and incident diabetes during the DPP

	Sample *N*	Weighted *N*	Without adjustment		With adjustment	
			HR (95% CI)	*P*	HR (95% CI)	*P*
Isoleucine	741	3,180	1.24 (1.03, 1.50)	0.03	1.14 (0.95, 1.36)	0.15
Leucine	740	3,179	1.27 (1.05, 1.53)	0.01	1.15 (0.97, 1.37)	0.11
Valine	741	3,180	1.22 (1.02, 1.47)	0.03	1.13 (0.96, 1.34)	0.14
Phenylalanine	741	3,180	1.03 (0.89, 1.19)	0.69	0.98 (0.85, 1.12)	0.72
Tyrosine	741	3,180	1.16 (1.02, 1.33)	0.03	1.09 (0.93, 1.27)	0.25
Glutamine/glutamate	741	3,180	0.82 (0.72, 0.93)	0.002	0.88 (0.74, 1.04)	0.13

HR is diabetes risk for each SD increase in log metabolite level. Model results are shown without and with adjustment for age, sex, BMI, FPG, hypertension (yes/no), race/ethnicity, and treatment group. Sample *N* indicates number of measured values and weighted *N* indicates number of samples included in the IPW cohort. Results are from the IPW cohort.

phenylalanine 0.95 [0.51, 1.76], *P* = 0.87; tyrosine 1.12 [0.73, 1.73], *P* = 0.61; and glutamine/glutamate 0.85 [0.70, 1.04], *P* = 0.11). However, as in the IPW analyses, we found associations in the matched case-control cohort between incident diabetes and betaine (0.42 [0.21, 0.84], *P* = 0.01), serine (0.52 [0.30, 0.90], *P* = 0.01), and propionylcarnitine (1.40 [0.99, 1.98], *P* = 0.05). The association of methionine sulfoxide with incident diabetes in the matched case-control cohort was in the same direction as in the IPW analyses, but it did not reach statistical significance (1.23 [0.82, 1.48], *P* = 0.31). Thus, analyses performed using the traditional approach were consistent with analyses using the IPW approach.

A significant interaction between baseline metabolite levels and treatment group was detected in the statistical models for several metabolites, indicating that the relationship between incident diabetes and the metabolite level differed across the intensive lifestyle, metformin, and placebo groups (Supplementary Table 2). Notably, the association of baseline methionine sulfoxide, betaine, serine, and propionylcarnitine with incident diabetes, which predicted diabetes incidence in the DPP, did not differ across the intensive lifestyle, metformin, or placebo arms.

To assess whether interventions directly influenced metabolite levels, we compared the metabolite change from baseline to 2 years of treatment during the DPP between participants assigned to the placebo, metformin, and intensive lifestyle arms (Table 4). None of the previously established amino acid markers of diabetes

risk (BCAA/As or glutamine/glutamate) were significantly altered by the intensive lifestyle or metformin intervention as compared with placebo. However, four metabolites were altered by preventive interventions (Table 4). Betaine levels were increased by the intensive lifestyle intervention and were nominally higher in the metformin group as compared with the placebo group. Notably, increases in betaine at 2 years of intervention were associated with lower incident diabetes in the DPP (HR 0.82 [95% CI 0.71, 0.95], *P* = 0.008) (Supplementary Table 3).

The relationship between incident diabetes and the metabolite change at 2 years of intervention differed across the intensive lifestyle, metformin, and placebo groups for several metabolites in the DPP (Supplementary Table 4). As with the association of baseline metabolite levels and incident diabetes, the association of change in methionine sulfoxide, betaine, serine, and propionylcarnitine with incident diabetes did not differ across the intensive lifestyle, metformin, or placebo arms.

DISCUSSION

We report the novel characterization of the metabolite profile associated with diabetes incidence in the DPP, a clinical trial for diabetes prevention in a high-risk multiethnic population. By measuring plasma metabolites before and during lifestyle, metformin, and placebo treatments, we find that increased baseline levels of BCAA/As and glutamine/glutamate, which have previously been shown to predict type 2 diabetes in populations at average risk, are

Table 3—Association of metabolites with incident diabetes during the DPP

	Sample *N*	Weighted *N*	HR (95% CI)	*P*
Methionine sulfoxide	741	3,180	1.12 (1.04, 1.22)	0.006
Betaine	740	3,179	0.84 (0.73, 0.97)	0.02
Serine	741	3,180	0.87 (0.76, 0.99)	0.04
Propionylcarnitine	741	3,180	1.20 (1.01, 1.44)	0.04

HR is diabetes risk for each SD increase in log metabolite level. Metabolites are shown in order of ascending *P* value. Model results are adjusted for age, sex, BMI, FPG, hypertension (yes/no), race/ethnicity, and treatment group. Sample *N* indicates number of measured values and weighted *N* indicates number of samples included in the IPW cohort. Results are from the IPW cohort.

Figure 1—Correlation matrix of BCAA/As, glutamine/glutamate, methionine sulfoxide, betaine, serine, propionylcarnitine, and baseline metabolic traits in the DPP. *A*: Correlation coefficients. *B*: *P* values.

also associated with diabetes incidence in the DPP. Although these relationships are attenuated and become nonsignificant after adjustment for clinical characteristics, several other metabolites, including betaine, serine, and methionine sulfoxide, remain associated with incident diabetes during the DPP. Together, our results suggest that a distinct metabolite profile, including metabolites within the choline metabolic pathway, is informative with

respect to future type 2 diabetes incidence in a high-risk population. Further, our work in humans complements emerging animal data demonstrating a direct role for dietary betaine in promoting metabolic health, as shown by Ejaz et al. (14). This study provides a potential mechanism by which betaine improves glucose tolerance, bolstering the association of betaine with reduced incident diabetes that is demonstrated in our independent work. A brief summary of this work is provided in the Supplementary Data.

The strengths of the current study's design include the well-phenotyped, multiethnic cohort; the randomized allocation of lifestyle or metformin interventions; the inclusion of a placebo group; and the standardized ascertainment of metabolite levels at baseline and follow-up using a well-established platform. The participants with and without incident diabetes on whom metabolite profiling was performed were closely matched, and participants free of diabetes were included at multiple time points; thus, our sampling design increased the likelihood of identifying metabolites that provided nonoverlapping information with clinical measures. Although sample availability limited metabolite profiling to a subset of samples from the DPP, the main treatment effects of the DPP were restored after implementation of IPW, and demographic characteristics mirrored the same in the original DPP cohort.

As with prior reports (2,3), we find that higher levels of BCAA/As and lower levels of glutamine/glutamate are associated with greater likelihood of incident diabetes. The direction of these effects is the same as that reported previously (2). However, these effects did not remain significant after adjustment for clinical measures, suggesting that elevated BCAA/As or lower glutamine/glutamate may reflect the combined effects of other risk factors such as obesity, elevated fasting glucose, and impaired glucose tolerance in this high-risk population. BCAA/As and glutamine/glutamate were also not modulated by preventive interventions for diabetes, and the change in these metabolites was not associated with future diabetes development.

Together, these results suggest that these amino acids may not contribute to the biological transition from impaired glucose metabolism to type 2 diabetes. These findings do not undermine previously reported associations of BCAA/As and glutamine/glutamate with incident diabetes in observational cohorts in populations at average risk of diabetes (2–4). Rather, they support that these established markers may be stronger predictors of incident diabetes prior to the emergence of a high-risk metabolic state common among DPP participants. Further, diabetes risk biomarkers may be different in high-risk individuals entering clinical trials compared with individuals in community-based cohorts. Differences between the DPP and population-based cohorts in BMI, fasting or postprandial glucose and insulin, or activity level may explain these differences. These hypotheses can be tested in further experiments. We call attention to prior work from our group and others demonstrating that BCAA/As change with the restoration of insulin sensitivity following bariatric surgery (15) and are altered by acute interventions that modulate insulin action (16). Thus, BCAA/As correlate with metabolic states in other clinical settings.

Our study identified the choline metabolites betaine, serine, and methionine sulfoxide as associated with incident diabetes even after adjustment for clinical measures in the DPP. The relationship between baseline levels of these metabolites and incident diabetes did not differ among DPP participants randomized to placebo, metformin, or lifestyle arms, suggesting that these markers could be informative irrespective of whether an individual does or does not engage in preventive interventions for diabetes. These relationships were also observed in the traditional conditional logistic approach using only measured metabolite data. Thus, choline metabolites may be markers of future type 2 diabetes that merit additional validation in heterogeneous cohorts, potentially among the individuals in high-risk groups.

Betaine is of specific interest for several reasons. First, prior cross-sectional analyses, including our own work,

Table 4—Percent change in metabolites from baseline by DPP treatment group

	Lifestyle	Metformin	Placebo	*P* (lifestyle vs. placebo)	*P* (metformin vs. placebo)
Changes in established amino acid markers of diabetes					
Isoleucine	12.1	20.9	9.4	0.70	0.27
Leucine	9.9	15.0	7.2	0.61	0.34
Valine	7.9	9.9	4.8	0.62	0.47
Phenylalanine	8.9	11.6	10.2	0.84	0.87
Tyrosine	9.8	12.0	11.5	0.83	0.96
Glutamine/glutamate	49.9	91.8	71.7	0.49	0.64
Changes in additional metabolites					
Betaine‡	12.8	7.9	5.4	0.01	0.43
Allantoin	21.2	8.2	4.4	0.03	0.49
1-Methylhistamine	20.3	11.7	26.2	0.34	0.01
Hexacosanoyl carnitine	4.5	7.7	20.3	0.02	0.12

Data are percent change at 2 years of intervention from baseline in the DPP for each treatment group. Metabolite changes for established amino acid markers of diabetes and metabolite changes for any comparison at *P* < 0.05 are presented. ‡Comparison of lifestyle vs. metformin change for betaine is significant at *P* = 0.04.

have shown that high levels of betaine are associated with other favorable metabolic traits (4,17,18). Notably, Ejaz et al. (14) report that betaine was the top-ranking metabolite associated with insulin resistance in normoglycemic humans. Second, our current prospective study of DPP patient samples demonstrates that betaine levels were increased at 2 years by the intensive lifestyle intervention and numerically raised by metformin treatment, interventions that were successful in reducing diabetes incidence (1). Finally, the increase in betaine at 2 years of intervention was associated with reduced diabetes incidence at the end of the DPP among all participants. Therefore, betaine might have utility in the monitoring for the effect of preventive interventions on future diabetes development. For example, individuals who do not have increases in betaine during lifestyle interventions may require closer monitoring or additional interventions to prevent the onset of type 2 diabetes. These clinically relevant hypotheses require validation in future studies.

Betaine is an osmolyte and methyl donor, and systemic betaine is derived either from choline or obtained directly from the diet. Indeed, only approximately 20% of the variation in circulating betaine levels is influenced by host genetic factors (19), and these genetic variants have not been associated with glycemic traits or type 2 diabetes (18). Collectively, these data suggest a dominant role for dietary intake or other environmental factors in influencing plasma levels. In mouse models, dietary betaine supplementation not only increases plasma betaine levels but also improves multiple aspects of metabolic health, reducing hepatosteatosis through fibroblast growth factor 21–mediated pathways (14). These data support the hypothesis that betaine is not only a marker but also a modifier of improved systemic metabolism. Gut flora have an important role in the generation of metabolites from dietary choline (20), including organic amines such as betaine, which are associated with favorable metabolic phenotypes (4,17,18), and trimethylamine-N-oxidase (TMAO), which has been associated with increased cardiovascular disease (20–22). Indeed, under the influence of the microbiome, TMAO can be generated from dietary choline or, indirectly, from betaine. Still, it remains unclear whether TMAO is a direct marker of cardiovascular disease or is confounded by other disease states (23–25). Thus, the interaction between the microbiome and other host factors, dietary intake, and metabolite biomarkers needs to be clarified in preclinical studies, such as those provided by Ejaz et al. (14), and in early phase clinical studies prior to implementing long-term trials that test the effect of dietary betaine supplementation in individuals at high risk of diabetes.

The limitations of our study deserve comment. First, we measured a subset of 84 circulating metabolites out of the estimated 5,000–10,000 species within the metabolome. Our results are thus agnostic to the predictive capacity of metabolites not measured on our platform (26,27). Utilization of novel, nontargeted measures of metabolites may highlight additional markers of diabetes risk within this population. Although our method is robust to the measurement of small organic amines, alternative methods of metabolite profiling, specifically nuclear magnetic resonance, could be used to extend our findings. Second, samples for direct measurement were selected based on a nested case-control design and established statistical methods were used to extend the results to the entire DPP. This approach was chosen given logistical issues associated with sample availability and mirrored that applied in other studies (2,28). Traditional analyses, using conditional logistic regression on measured data that confirmed the relationships of serine and betaine with incident diabetes, are reassuring for the robustness of the novel results. Last, our results will require confirmation. The unique aspects of the DPP (high-risk multiethnic population, randomized to preventive treatments, long-term follow-up, and sample available for metabolite profiling) make validation of our findings within a similar cohort challenging at present. Nevertheless, our observations in a longitudinal human cohort provide strong support for a relationship between betaine and diabetes risk and implicate choline metabolic intermediates as contributors to the pathogenesis of insulin resistance and diabetes risk. Human intervention studies will be required to fully test these hypotheses.

In summary, metabolite profiling in the DPP demonstrates a unique signature of diabetes risk. Specifically, associations of previously validated amino acid markers with diabetes risk are present, but they are more modest than in population-based cohorts and become attenuated after consideration of clinical measures. By contrast, new markers, including betaine and serine, emerge as indicators of diabetes risk prior to and during preventive interventions. Further studies are needed to formally test the capacity of these new metabolites as markers of diabetes incidence in both low- and high-risk populations and to test whether interventions modulating these metabolites may have beneficial impacts on metabolism and reducing diabetes risk.

References

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

2. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. Nat Med 2011;17:448–453

3. 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

4. Cheng S, Rhee EP, Larson MG, et al. Metabolite profiling identifies pathways associated with metabolic risk in humans. Circulation 2012;125:2222–2231

5. 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

6. DPP. The Diabetes Prevention Program. Design and methods for a clinical trial in the prevention of type 2 diabetes. Diabetes Care 1999;22:623–634

7. Lewis GD, Wei R, Liu E, et al. Metabolite profiling of blood from individuals undergoing planned myocardial infarction reveals early markers of myocardial injury. J Clin Invest 2008;118:3503–3512

8. 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

9. Roberts LD, Souza AL, Gerszten RE, Clish CB. Targeted metabolomics. Curr Protoc Mol Biol 2012;Chapter 30:Unit 30.2.1–24

10. The Diabetes Prevention Program Research Group. The Diabetes Prevention Program: baseline characteristics of the randomized cohort. Diabetes Care 2000; 23:1619–1629

11. American Diabetes Association. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 1997;20:1183–1197

12. Samuelsen SO. A psudolikelihood approach to analysis of nested case-control studies. Biometrika 1997;84:379–394

13. Støer NC, Samuelsen SO. Comparison of estimators in nested case-control studies with multiple outcomes. Lifetime Data Anal 2012;18:261–283

14. Ejaz A, Martinez-Guino L, Goldfine AB, et al. Dietary betaine supplementation increases Fgf21 levels to improve glucose homeostasis and reduce hepatic lipid accumulation in mice. Diabetes 2016;65:902–912

15. Magkos F, Bradley D, Schweitzer GG, et al. Effect of Roux-en-Y gastric bypass and laparoscopic adjustable gastric banding on branched-chain amino acid metabolism. Diabetes 2013;62:2757–2761

16. Walford GA, Davis J, Warner AS, et al. Branched chain and aromatic amino acids change acutely following two medical therapies for type 2 diabetes mellitus. Metabolism 2013;62:1772–1778

17. Konstantinova SV, Tell GS, Vollset SE, Nygård O, Bleie Ø, Ueland PM. Divergent associations of plasma choline and betaine with components of metabolic syndrome in middle age and elderly men and women. J Nutr 2008;138: 914–920

18. Xie W, Wood AR, Lyssenko V, et al.; MAGIC Investigators; DIAGRAM Consortium; GENESIS Consortium; RISC Consortium. Genetic variants associated with glycine metabolism and their role in insulin sensitivity and type 2 diabetes. Diabetes 2013;62:2141–2150

19. Rhee EP, Ho JE, Chen MH, et al. A genome-wide association study of the human metabolome in a community-based cohort. Cell Metab 2013;18:130–143

20. Tang WH, Wang Z, Levison BS, et al. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. N Engl J Med 2013;368:1575–1584

21. Tang WH, Wang Z, Fan Y, et al. Prognostic value of elevated levels of intestinal microbe-generated metabolite trimethylamine-N-oxide in patients with heart failure: refining the gut hypothesis. J Am Coll Cardiol 2014;64:1908–1914

22. McCarty MF, DiNicolantonio JJ. Trimethylamine-N-oxide and heart failure. J Am Coll Cardiol 2015;66:94–96

23. McEntyre CJ, Lever M, Chambers ST, et al. Variation of betaine, N,N-dimethylglycine, choline, glycerophosphorylcholine, taurine and trimethylamine-N-oxide in the plasma and urine of overweight people with type 2 diabetes over a two-year period. Ann Clin Biochem 2015;52:352–360

24. Mueller DM, Allenspach M, Othman A, et al. Plasma levels of trimethylamine-N-oxide are confounded by impaired kidney function and poor metabolic control. Atherosclerosis 2015;243:638–644

25. Rhee EP, Souza A, Farrell L, et al. Metabolite profiling identifies markers of uremia. J Am Soc Nephrol 2010;21:1041–1051

26. Gerszten RE, Wang TJ. The search for new cardiovascular biomarkers. Nature 2008;451:949–952

27. Rhee EP, Gerszten RE. Metabolomics and cardiovascular biomarker discovery. Clin Chem 2012;58:139–147

28. Walford GA, Ma Y, Christophi CA, et al.; Diabetes Prevention Program Research Group. Circulating natriuretic peptide concentrations reflect changes in insulin sensitivity over time in the Diabetes Prevention Program. Diabetologia 2014;57:935–939