Published in final edited form as:

Metabolism. 2016 September ; 65(9): 1399–1408. doi:10.1016/j.metabol.2016.06.007.

Validation of a metabolite panel for early diagnosis of type 2 diabetes

Tonia C. Carter^a, Dietrich Rein^b, Inken Padberg^c, Erik Peter^d, Ulrike Rennefahrt^d, Donna E. David^e, Valerie McManus^f, Elisha Stefanski^e, Silke Martin^g, Philipp Schatz^b, and Steven J. Schrodi^{a,*}

^aCenter for Human Genetics, Marshfield Clinic Research Foundation, 1000 North Oak Avenue, Marshfield, Wisconsin, 54449, USA

^bMetanomics Health GmbH, Tegeler Weg 33, 10589 Berlin, Germany

^cCenter for Stroke Research Berlin, Berlin, Germany

^dmetanomics GmbH, Tegeler Weg 33, 10589 Berlin, Germany

^eIntegrated Research and Development Laboratory, Marshfield Clinic Research Foundation, 1000 North Oak Avenue, Marshfield, Wisconsin, 54449, USA

^fBiomedical Informatics Research Center, Marshfield Clinic Research Foundation, 1000 North Oak Avenue, Marshfield, Wisconsin, 54449, USA

^gBlutspendedienst des Bayerischen Roten Kreuzes gemeinnützige GmbH, München, Germany

Abstract

Background—Accurate, early diagnosis of type 2 diabetes (T2D) would enable more effective clinical management and a reduction in T2D complications. Therefore, we sought to identify plasma metabolite and protein biomarkers that, in combination with glucose, can better predict future T2D compared with glucose alone.

Methods—In this case-control study, we used plasma samples from the Bavarian Red Cross Blood Transfusion Center study (61 T2D cases and 78 non-diabetic controls) for discovering T2D-associated metabolites, and plasma samples from the Personalized Medicine Research Project in

* **Correspondence to:** Steven Schrodi, PhD, Center for Human Genetics, MLR, Marshfield Clinic, 1000 North Oak Avenue, Marshfield, WI 54449, Tel: 715-221-6443, Fax: 715-389-4950, schrodi.steven@mcrf.mfldclin.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflict Of Interest

T.C.C., D.E.D., V.M., E.S., S.M., and S.J.S. declare no conflicts of interest. D.R. and P.S. are employed by Metanomics Health GmbH. I.P. was employed by metanomics GmbH during conduct of the study. E.P. and U.R. are employed by metanomics GmbH. The study was partially funded by Metanomics Health GmbH.

Author Contributions

D.R. and P.S. conceived the study. D.R., P.S., and S.J.S. contributed to the design of the study. D.R., I.P., D.E.D., V.M., E.S., S.M. collected data, and D.R., I.P., E.P., U.R., S.J.S., and T.C.C. performed data analysis and interpretation. T.C.C. drafted the manuscript and all authors contributed to editing and critical review of the manuscript. All authors have reviewed and approved the final version of the manuscript.

Wisconsin (56 T2D cases and 445 non-diabetic controls) for validation. All samples were obtained before or at T2D diagnosis. We tested whether the T2D-associated metabolites could distinguish incident T2D cases from controls, as measured by the area under the receiver operating characteristic curve (AUC). Additionally, we tested six metabolic/pro-inflammatory proteins for their potential to augment the ability of the metabolites to distinguish cases from controls.

Results—A panel of 10 metabolites discriminated better between T2D cases and controls than glucose alone (AUCs: 0.90 vs 0.87; $p = 2.08 \times 10^{-5}$) in Bavarian samples, and associations between these metabolites and T2D were confirmed in Wisconsin samples. With use of either a Bayesian network classifier or ridge logistic regression, the metabolites, with or without the proteins, discriminated incident T2D cases from controls marginally better than glucose in the Wisconsin samples, although the difference in AUCs was not statistically significant. However, when the metabolites and proteins were added to two previously reported T2D prediction models, the AUCs were higher than those of each prediction model alone (AUCs: 0.92 vs 0.87; $p = 3.96 \times 10^{-2}$ and AUCs: 0.91 vs 0.71; $p = 1.03 \times 10^{-5}$, for each model, respectively).

Conclusions—Compared with glucose alone or with previously described T2D prediction models, a panel of plasma biomarkers showed promise for improved discrimination of incident T2D, but more investigation is needed to develop an early diagnostic marker.

Keywords

biomarker; diabetes; metabolomics; glucose; insulin resistance; predictive modeling

1. Introduction

Type 2 diabetes (T2D) is a chronic, progressive metabolic disease that is a major cause of mortality and morbidity globally [1]. Through earlier and more accurate T2D diagnosis, improved T2D treatment and monitoring of T2D progression can be achieved. T2D is associated with shifts in the concentrations of small molecules, peptides, hormones, and enzyme activities, in a variety of tissues, including blood [2]. Given the substantially altered biochemical profile observed in T2D, molecular markers could provide the basis for early T2D diagnosis.

Advances in the application of metabolite profiling for predicting T2D risk have occurred in recent years. An example is the study by Floegel and colleagues who used a targeted metabolomics approach and demonstrated that a linear factor equation composed of amino acids, phosphatidylcholines, and hexoses was significantly correlated in a dose-response manner with T2D risk in two independent cohorts (EPIC-Potsdam and KORA) [3]. These authors also showed that a related set of metabolites moderately increased the area under the receiver operating characteristic curve (AUC) for T2D risk when added to a model of traditional risk factors: glucose and hemoglobin A1c (HbA1c). Other studies have applied metabolomics to prediabetes or incident T2D, showing, for example, the protective effects of glycine and lysophosphatidylcholine C18:2 on risk of T2D onset, with reasonable predictive performance [4].

To identify plasma metabolites that can potentially improve prediction of future T2D, compared with prediction based on plasma glucose alone, we used a non-targeted metabolomics approach for discovery of metabolic alterations associated with T2D onset. The T2D-associated metabolites were then validated in an independent cohort. We also assessed whether including proteins related to metabolic syndrome and/or pro-inflammatory responses led to improvement in T2D risk classification. Biomarkers that are associated with prediabetes or T2D onset are potentially applicable for diagnosis of therapy-naïve persons with T2D, to test for association with T2D in biobank samples, and as a tool to improve T2D early-intervention therapeutic strategies.

2. Methods

2.1. Study design

To identify a robust metabolite panel for early T2D diagnosis, we conducted a case-control study with two independent study populations: one for biomarker identification (discovery group) and the other for confirmatory testing (validation group). For biomarker discovery, we used fasting blood samples from the Bavarian Red Cross Blood Transfusion Center (BRC) study, described previously in Padberg et al. [5]. Blood samples were collected at the time of first T2D diagnosis. For validation, we used data and blood samples from the Personalized Medicine Research Project (PMRP) in Marshfield, Wisconsin, USA [6]. Blood samples were collected at PMRP enrollment which was within 18 months before the first T2D diagnosis. Therefore, in the validation group, we examined whether the biomarkers could predict a future T2D diagnosis within 18 months.

2.2. Study subjects

In the discovery group, cases were 61 BRC study subjects who had a high FindRisk score (indicating high T2D risk) after completing a FindRisk questionnaire [7] and who provided a blood sample that tested positive for T2D. This led to a first-time T2D diagnosis at the time of BRC study enrollment. Seventy-eight non-diabetic BRC study subjects who were matched to cases by age, sex, and body mass index (BMI) served as controls in the discovery group.

The PMRP is a population-based biobank that links stored biological samples to longitudinal electronic medical records from the Marshfield Clinic for about 20,000 central Wisconsin adults [6]. Approximately 98% of participants are of European ancestry with > 70% claiming German ancestry. Fifty-six PMRP participants diagnosed with T2D for the first time within 18 months after PMRP enrollment (cases), and 445 non-diabetic PMRP participants (controls) matched to cases by age, sex, and BMI, formed the validation group. For the 56 cases, the median (inter-quartile range) time to diabetes diagnosis after PMRP enrollment was 275 (139–417.5) days. The validation group had more females and a greater mean BMI than the discovery group (Table 1).

The procedure used for matching cases and controls in the discovery and validation groups is described in Supplementary Methods.

All subjects gave informed consent and the study was approved by the institutional review boards of the Bavarian Chamber of Physicians (Ethik-Kommission der Bayrischen Landesärztekammer) and the Marshfield Clinic Research Foundation.

2.3. Type 2 diabetes status

A T2D diagnosis in the discovery group was based on American Diabetes Association criteria: fasting plasma glucose (FPG) concentration ≥ 126 mg/dl and/or plasma glucose concentration 2 h after a standardized 75 g oral glucose tolerance test (OGTT) ≥ 200 mg/dl and/or HbA1c $\geq 6.5\%$. A subject was considered to be diabetic if at least one of their blood samples met these criteria and tests of at least one additional blood sample indicated the subject to be diabetic or at risk for diabetes. In addition, about 50% of the so identified diabetic subjects were questioned three years later regarding their diabetes status. If a subject did not confirm the diabetes status, he/she was excluded from the dataset. Controls in the biomarker development group had values ≤ 100 mg/dl for fasting plasma glucose and ≤ 140 mg/dl after the OGTT.

A validated algorithm for identifying patients with T2D from electronic medical record data [8] was used to determine a T2D diagnosis in the validation group. The diagnosis was based on the presence of at least two International Classification of Diseases – 9th Revision – Clinical Modification (ICD-9-CM) codes for T2D (indicating a doctor's diagnosis of T2D) along with intake of diabetes medication and/or an abnormal laboratory value (FPG > 125 mg/dL, random plasma glucose > 200 mg/dL, and/or HbA1c $\geq 6.5\%$) within 18 months after PMRP enrollment. The electronic medical records of controls in the validation group had no ICD-9-CM codes for T2D and all measurements of FPG, random plasma glucose and HbA1c were within the normal range. Thus, the discovery and validation groups were expected to be comparable as T2D in the validation group was mostly diagnosed by a clinician based on the results of FPG, OGTT, and HbA1c tests. Table 1 shows basic descriptive characteristics of the two study groups.

2.4. Blood samples

Blood samples were collected from fasted subjects (≥ 8 h since last meal). Details of blood sample processing are described in Supplementary Methods.

2.5. Metabolite profiling

Metabolite profiling of all plasma samples was performed by metanomics GmbH (Berlin, Germany). Two types of mass spectrometry analyses were used for relative quantification of metabolites, as described in Supplementary Methods. All metabolite measurements were performed blinded to case-control status.

2.6. Metabolite data normalization

Metabolite profiling generated semi-quantitative measurements of metabolite concentrations, that is, the concentration of a metabolite in each study sample relative to the concentration of the metabolite in a reference sample. We used separate reference samples for the discovery and validation groups, and each reference sample consisted of a pool of combined aliquots of all plasma samples specific to either the discovery or validation group.

Following comprehensive analytical validation steps, data were normalized to the median of the reference samples to account for inter- and intra-instrumental variation.

2.7. Protein quantification

Seven metabolic disease-related proteins were selected for quantification in plasma samples from the validation group to determine whether the addition of these proteins would improve the ability of metabolite panels to predict a T2D diagnosis. Proteins were chosen based on their established correlations with T2D status in the literature and on the availability of existing protein quantitation assays (multiplex ELISAs) from Meso-Scale Discovery (Rockville, Maryland, USA) using their SECTOR Imager 2400 platform: C-peptide, adiponectin, interleukin 6 (IL-6), active glucagon-like peptide-1 (GLP-1), insulin, glucagon, and leptin. Details of protein quantification assays are described in Supplementary Methods. All protein measurements were acquired blinded to case-control status. Because the plasma samples were collected without aprotinin, we were unable to accurately measure active GLP-1. Therefore, GLP-1 results were disregarded.

2.8. Statistical analysis

We performed a log10-transformation of the pool-normalized ratio for each metabolite, and of protein concentrations, to obtain an approximately normal distribution of the data. We checked data quality as described in Supplementary Methods. SIMCA-P version 10.0 (Umetrics AB, Umeå, Sweden), TIBCO® Spotfire® 3.3.1, R version 2.14.1 (packages *glmnet*, *randomForest* and *ROCR*), and WEKA [9] were used for data analyses and visualizations.

2.8.1. Discovery of metabolite panels for T2D—To identify metabolite combinations that might improve T2D diagnosis compared with plasma glucose alone, we applied classification tools (penalized logistic regression, random forest, and elastic net approaches) to metabolite profiling data obtained for the discovery group. A maximum of 10 metabolites at a time were used for classification because testing for a larger number of metabolites is likely to be impractical in clinical practice. In addition, a fixed panel size avoids overfitting due to picking a seemingly optimal number of metabolites for a particular dataset, and our preliminary analyses indicated that increasing the number of metabolites in a panel did not substantially improve the performance of the panel to separate T2D cases from controls. We ranked each panel of up to 10 metabolites on its ability to distinguish between T2D cases and controls using a feature selection approach that had the following characteristics: (1) forward selection based on *t*-test result (accepts correlation between metabolites), (2) forward selection maximizing orthogonality (minimizes correlation between metabolites), (3) recursive feature elimination, and (4) elastic net. The approach was based upon training an elastic net logistic regression model (starting with intercept only), selecting an additional metabolite that had the highest absolute correlation with the residuals of the model, and iteration of the selection of additional metabolites until the desired number of features was reached. We used this approach in the expectation that it would lead to adding metabolites that contributed the maximum orthogonal information to the model. The embedding of this feature selection strategy in 10-fold cross validation allowed us to obtain realistic performance estimates of each metabolite panel as a classifier of case-control status in the

discovery group. We used AUCs to assess the performance of metabolite panels, compared with glucose only, to distinguish persons diagnosed with T2D from controls. The corrected re-sampled *t*-test [10] was used to compare AUCs.

Because the validation of T2D-associated metabolite panels (identified using the discovery group) was not possible if one or more of the metabolites in the panels could not be measured in the validation group (possibly due to sample quality differences between the two study groups), data for the validation group were checked for the presence of all metabolites in panels associated with T2D. If any metabolite in a panel was absent from the validation data, a new round of analysis was performed using data for the discovery group to identify another panel of T2D-associated metabolites.

2.8.2. Validation of metabolite associations with T2D—For panels that had an AUC greater than that of glucose alone, we further examined the ability of the metabolites in these panels to discriminate between T2D cases and controls by performing analysis of variance (ANOVA) to compare metabolite pool-normalized ratios between T2D cases and controls in the discovery group. We adjusted the ANOVA models for age at study (BRC study) enrollment, sex, BMI, interaction of sex and age, and interaction of sex and BMI. For validation, we performed ANOVA to compare pool-normalized ratios of the same metabolites between T2D cases and controls in the validation group. We corrected for multiple comparisons using the Bonferroni-Holm method.

2.8.3. Analysis of metabolites and proteins combined—We performed ANOVA to compare plasma protein concentrations between T2D cases and controls in the validation group, with adjustment for the same covariates as in the analysis of metabolite panels. We also calculated Pearson correlation coefficients to examine correlations between protein and metabolite concentrations. Next, we used a Bayesian network classifier [11], with 10-fold cross validation, to test the ability of (a) the combination of metabolites and all six proteins, (b) the metabolites only, and (c) glucose only, to discriminate between the incident T2D cases and non-diabetic controls in the validation group. We used the corrected re-sampled *t*-test [10] to compare AUCs between the biomarker groups. Each AUC was calculated by collecting all the predicted probabilities of being a T2D case from each fold of 10-fold cross validation. To determine whether another algorithm would produce results similar to those of the Bayesian network classifier, we repeated the comparisons using ridge logistic regression [12] with 10-fold cross validation.

We also used the Bayesian network and ridge logistic regression classifiers to examine the ability of two previously reported T2D prediction models [13, 14] to distinguish between cases and controls in the validation group, and to determine whether T2D prediction by these models could be improved by adding the T2D-associated metabolites we identified. We used electronic medical record data, existing genotype data, and metabolic profiling data for the validation group to generate variables for the models. For the model of Wilson et al. [13], we used family history of T2D in a first-degree relative as a substitute for parental history of T2D. We had no data for some of the variables in the Walford et al. [14] model (circulating concentrations of phosphatidylcholine C38:6, triacylglycerol C44:1,

triacylglycerol C48:0, triacylglycerol C52:1, and triacylglycerol C56:9); therefore, we did not include these variables in the model.

3. Results

Metabolite profiling identified 234 metabolites that surpassed quality criteria; 190 (81.2%) had a known structure (based on a database of metabolite structures maintained by metanomics, Berlin, Germany). Of the 234 metabolites, we elected to focus on the 168 (71.8%) (Supplementary Table 1), representing various metabolite classes (Supplementary Figure 1), that were detected in both the discovery and validation groups. Analysis of all metabolites showed that biological variability between T2D cases and non-diabetic controls exceeded the technical variability (occurring as a result of performing measurements on different instruments and different days) of the reference samples (pools). The technical variability, determined by the median relative standard deviation of pooled samples from the two study groups, was 7.6% and 8.3% in the discovery and validation groups, respectively. These values are an indicator of the low technical variability of the metabolite profiling method used. In contrast, the median within-group variability (comprising both biological and technical variability) was higher than the technical variability alone for both T2D cases (22.0% in the discovery group and 27.1% in the validation group) and controls (21.5% in the discovery group and 25.9% in the validation group).

In the discovery and validation groups, separately, univariate ANOVA of metabolite profiling data revealed many previously reported differences in metabolite concentrations between T2D cases and controls (Supplementary Table 2). These changes in metabolite concentrations have been shown to predict T2D years in advance or to be associated with overt T2D [15]. This consistency between our findings and those reported by others is a validation of our metabolite profiling data. In addition to the expected increase in glucose concentration, our dataset confirmed increases in branched-chain amino acids (isoleucine, leucine, and valine) [16], tyrosine (aromatic amino acid) [16], 2-hydroxybutyrate [17], and mannose [18], as well as decreases in glycine [4], lysophosphatidylcholine C18:2 [4], and 1,5-anhydrosorbitol [19] in T2D. However, in contrast to previous reports, we observed no statistically significant differences in the concentrations of phenylalanine [20] and the ketone, 3-hydroxybutyrate [20], between T2D cases and controls.

We identified seven panels of multiple metabolites associated with T2D in the discovery group and trained either elastic net or random forest classifiers using the discovery data. Three of the seven panels had an AUC greater than that of plasma glucose alone ($AUC_{\text{glucose}} = 0.866$) for distinguishing T2D cases from controls (Table 2), suggesting that they might outperform glucose for discriminating between T2D cases and controls. Of the three (panels 5, 6, and 7), we focused on panel 7 because it had the most significant p value in the comparison of AUCs with plasma glucose alone (Table 2). Of the 10 metabolites in panel 7, seven (including glucose) showed statistically significant associations ($p < 0.05$) with T2D in both the discovery and validation groups, based on ANOVA modeling (Table 3). Of the seven, the metabolites most strongly associated with T2D (based on p value) were glucose, mannose, and glucosamine, while the metabolites with the largest increase and decrease in concentration between T2D cases and controls were the bile acid, taurochenodeoxycholic

acid, and cholesteryl ester C18:1, respectively (Table 3). T2D associations with metabolites on panels 5 and 6 are shown in Supplementary Table 3.

The AUC for panel 7 was 3% greater than that of plasma glucose alone (Table 2). In an attempt to improve diagnostic utility of the panel, we considered selected metabolic disease-related proteins as additional biomarkers for discriminating between incident T2D cases and controls. We measured the plasma concentrations of these proteins in the validation group and observed that five of the six proteins measured were significantly associated ($p < 0.05$) with incident T2D in ANOVA models (Table 4). Correlations between the concentrations of the proteins and the 234 metabolites identified in the study were low to moderate in T2D cases and controls (Supplementary Figures 2 and 3, and Supplementary Methods). With the use of a Bayesian network classifier in WEKA [9], the group of metabolites in panel 7, either alone or in combination with the six proteins, had a slightly larger AUC than glucose alone for distinguishing between incident T2D cases and controls in the validation group (Fig. 1A; Table 5). However, no statistically significant differences in AUCs were observed in paired comparisons of glucose alone and either the panel 7 metabolites alone (0.88 vs 0.89; $p = 0.65$) or the panel 7 metabolites with the proteins (0.88 vs 0.91; $p = 0.19$). Similar results were obtained using ridge logistic regression in WEKA (Fig. 1B; Table 5). For each paired comparison, the AUCs from ridge logistic regression were 0.88 vs 0.90 ($p = 0.24$) and 0.88 vs 0.89 ($p = 0.40$), respectively.

To characterize metabolite panel 7 further, we assessed whether the panel, with glucose removed, could predict T2D in the validation group. We also examined whether the panel (with glucose included) could improve the ability of previously described T2D prediction models [13, 14] to predict T2D in the validation group. When glucose was excluded, the AUCs for metabolite panel 7 were marginally lower than, but not significantly different from, the AUCs for the full panel and for glucose alone (Table 5). Similar results were obtained when metabolite panel 7 was combined with the six metabolic disease-related proteins. The combination of the Wilson et al. [13] prediction model (that included clinical characteristics and fasting blood glucose concentration) and metabolite panel 7, with or without the six proteins, had slightly larger AUCs than the Wilson et al. model alone. When a Bayesian network classifier was used, the increase in AUC after adding metabolite panel 7 and the six proteins to the Wilson et al. model was statistically significant (Table 5). In the Walford et al. [14] model, predictive factors were selected metabolites (including one panel 7 metabolite, lysophosphatidylcholine C18:2) and a genetic risk score based on 62 single nucleotide polymorphisms that have been associated with T2D, but not glucose. The addition of metabolite panel 7, with or without the six proteins, to the Walford et al. model resulted in a statistically significant increase in the AUC over that of the Walford et al. model alone (Table 5). The AUCs obtained after combining the models of Wilson et al. and Walford et al. and metabolite panel 7, with or without the six proteins, did not differ significantly from the AUCs of glucose alone (Table 5).

4. Discussion

A profile of the concentrations of multiple biochemical molecules is likely to have an advantage over a single biomarker for detecting incident T2D because the dysfunction of

many metabolic pathways precedes the development of T2D. We observed that a panel of 10 metabolites showed a small but statistically significant improvement in discrimination of T2D cases and controls, compared with glucose alone, and we validated this panel in an independent group of study subjects that had characteristics (European ancestry, age, body mass index) in common with the first study group. The group of 10 metabolites, either alone or in combination with six metabolic disease-related proteins, showed marginally improved ability to discriminate between incident T2D cases and non-diabetic controls compared with glucose alone. The improvement was observed using two different classification algorithms (Bayesian network classifier and ridge logistic regression) but was not statistically significant. However, when compared with two previously reported T2D prediction models, the panel of 10 metabolites and the six proteins significantly improved the performance of these models to predict T2D. Thus, the T2D-associated metabolites in this study could have the potential for application to T2D prediction.

Other reports provide support for the involvement of these metabolites and proteins in T2D. Circulating concentrations of mannose [18], 2-hydroxybutyrate [21], glyoxylate [5, 22], insulin [23], C-peptide [24], glucagon [25], and two bile acids, glycochenodeoxycholic acid and taurochenodeoxycholic acid [26], are increased in subjects with T2D compared with healthy subjects. Lower circulating concentrations of lysophosphatidylcholine C18:2 [3, 4], lignoceric acid (C24:0) [27], adiponectin [28], and C18:1n7 [29], a fatty acid found in cholesteryl esters, are associated with T2D risk. The pro-inflammatory cytokine, IL-6, interferes with insulin signaling in human adipocytes *in vitro*, and is overexpressed in adipocytes from individuals who are insulin resistant [30]. Serum concentrations of leptin, a hormone that regulates appetite and energy expenditure in low leptin states [31], are lower in normal-weight than obese subjects [32], and responsiveness to exogenous leptin is diminished in obese persons [33]. However, the role of one metabolite, glucosamine, in T2D in humans is unclear. Glucosamine infusion in nondiabetic rats decreases insulin-mediated glucose uptake and glycogen synthesis in skeletal muscle [34], but glucosamine has shown no effect on insulin-mediated metabolism in humans [35].

Some of these metabolites are also associated with insulin resistance in non-diabetic individuals. Higher fasting plasma 2-hydroxybutyrate is associated with insulin resistance in non-diabetic subjects independently of age, sex, and BMI, based on use of the euglycemic insulin clamp [17]. Plasma lysophosphatidylcholine species overall are inversely associated with insulin resistance in overweight or obese subjects without overt T2D after adjustment for age, sex, systolic blood pressure and either BMI or waist:hip ratio [36], based on homeostasis model assessment – insulin resistance (HOMA-IR) [37] evaluation. Fasting plasma lignoceric acid is associated with lower insulin resistance in non-diabetic adults after adjustment for selected demographic and clinical risk factors, based on the use of HOMA-IR [38]. Fasting plasma adiponectin is inversely associated with insulin resistance in non-diabetic adults evaluated using the euglycemic insulin clamp [39] and HOMA-IR [40] methods. Therefore, one possible reason why the panel of metabolites had slightly higher AUCs for classifying incident T2D cases and controls in this study, compared with glucose alone, is because of the ability to detect insulin resistance, a part of the metabolic dysregulation that leads to T2D [23].

When glucose was removed from our panel of metabolites, the panel had an AUC that was not significantly different from that of glucose alone for distinguishing T2D cases from controls in the validation group. This indicated that glucose was not the only metabolite making a substantial contribution to the panel's ability to discriminate between T2D cases and controls, and also that the group of remaining metabolites could classify T2D cases and controls just as well as glucose. Further evidence of the value of the panel was shown by the statistically significant increase in AUCs when the metabolite panel (with glucose included) and the six metabolic disease-related proteins were added to either of two previously described T2D prediction models, compared with the AUCs of either prediction model alone. However, when the metabolite panel, six proteins, and two predictive models were combined, the AUCs were not significantly greater than the AUCs for glucose alone. This is probably because most T2D cases in the validation group were already hyperglycemic at the time of blood sample collection (within 18 months before T2D diagnosis), leading to fasting blood glucose being identified as the biomarker that had the strongest association with T2D status. Some metabolites (mannose, 2-hydroxybutyrate, and lysophosphatidylcholine C18:2) in our panel have been shown to predict T2D 3–10 years ahead [3, 4, 21, 41]. Prospective studies with a longer timeframe before T2D diagnosis than our study would be needed to determine whether the metabolite panel has the potential to predict future T2D before blood hyperglycemia manifests. If the panel is found to have little potential to predict T2D prior to overt hyperglycemia, its clinical utility for early detection of T2D would likely be low and at high cost, compared with testing for glucose only. Biomarkers that are not closely linked with glucose concentration, and that reliably demonstrate changes in concentration well in advance of blood hyperglycemia and T2D diagnosis, would be more useful for predicting future T2D.

The strengths of our study include the use of a non-targeted metabolomics approach that enabled detection of a wide range of metabolites and the use of a validation group that provided prospective data for correlating metabolite concentrations with future T2D diagnosis. Further, our confirmation of previous associations between metabolite concentrations and T2D, and our observation of similar findings in two independent study groups that had different T2D ascertainment methods increase confidence in the robustness of our results. Our study also had some limitations. We measured metabolite profiles at only one point in time in both study groups, but because T2D has a progressive onset, metabolite profiles at multiple time points before T2D diagnosis would be useful to identify new, early, diagnostic biomarkers of T2D. Also, we were able to control for some known T2D risk factors by case-control matching and the inclusion of covariates in ANOVA models, but there remains the potential for confounding by other factors such as diet and level of physical activity. Finally, the study population included mainly overweight, middle-aged persons of European ancestry, and more investigation is needed to determine whether our results can be replicated in other populations.

In summary, a panel of multiple metabolites (carbohydrates, ketone bodies, organic acids, lipids, and bile acids) showed a small improvement in ability to stratify incident T2D cases and controls, compared with glucose alone or with two previously described T2D prediction models. The panel could potentially have some useful application in early T2D diagnosis because, in the biomarker validation group, it was able to identify T2D patients up to 18

months before diagnosis. Further validation of these biomarkers is required and, rather like HbA1c, their physiological, pathological, and analytical properties should be elucidated so the limitations of their use as T2D biomarkers can be known. The mechanisms underlying the involvement of most of these biomarkers in T2D pathogenesis are poorly understood. Investigation of these mechanisms could provide additional insight into the early metabolic alterations of T2D and aid the discovery of novel preventative or therapeutic interventions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to acknowledge the considerable aid in the study from Dr. Murray Brilliant, Dr. Zhan Ye, and Dr. Jennifer Meece, administrative work from Marlene Stueland, and institutional review board paperwork from Terrie Kitchner. The authors would further like to acknowledge the critical contributions to project management by Dr. Hajo Schiewe, an employee of Metanomics Health during the study.

Funding

The work was funded by Metanomics Health GmbH and supported by a grant from the Bayrische Forschungsstiftung [grant AZ 773-707 (Study title: Testmethoden für die Früherkennung von Diabetes)]. The study was also supported by the National Institutes of Health (NIH) – Clinical and Translational Science Award (CTSA) program, previously funded through the National Center for Research Resources (NCRR) [grant 1UL1RR025011] and the National Center for Advancing Translational Sciences (NCATS) [grant 9U54TR000021], and now by the NCATS [grant UL1TR000427]. Additional support was provided by the National Institute of General Medical Sciences [grant U01HG004608 to Essentia Institute for Rural Health/Marshfield Clinic Research Foundation], Marshfield Clinic, philanthropic gifts in support of medical research at the Marshfield Clinic, and the Wisconsin Genomics Initiative. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Abbreviations

ANOVA	analysis of variance
AUC	area under the receiver operating characteristic curve
BMI	body mass index
BRC	Bavarian Red Cross Blood Transfusion Center study
ELISA	enzyme-linked immunosorbent assay
EPIC-Potsdam	European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study
FPG	fasting plasma glucose
GLP-1	glucagon-like peptide-1
HbA1c	hemoglobin A1c
HOMA-IR	homeostasis model assessment – insulin resistance
ICD-9-CM	International Classification of Diseases – 9 th Revision – Clinical Modification

IL-6	interleukin 6
KORA	Cooperative Health Research in the Region of Augsburg study
OGTT	oral glucose tolerance test
PMRP	Personalized Medicine Research Project
T2D	type 2 diabetes

References

1. van Dieren S, Beulens JW, van der Schouw YT, Grobbee DE, Neal B. The global burden of diabetes and its complications: an emerging pandemic. *Eur J Cardiovasc Prev Rehabil.* 2010; 17(Suppl 1):S3–8. [PubMed: 20489418]
2. Friedrich N. Metabolomics in diabetes research. *J Endocrinol.* 2012; 215:29–42. [PubMed: 22718433]
3. Floegel A, Stefan N, Yu Z, Mühlenbruch K, Drogan D, Joost HG, et al. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes.* 2013; 62:639–48. [PubMed: 23043162]
4. Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y, et al. Novel biomarkers for prediabetes identified by metabolomics. *Mol Syst Biol.* 2012; 8:615. [PubMed: 23010998]
5. Padberg I, Peter E, Gonzalez-Maldonado S, Witt H, Mueller M, Weis T, et al. A new metabolomic signature in type-2 diabetes mellitus and its pathophysiology. *PLoS One.* 2014; 9:e85082. [PubMed: 24465478]
6. McCarty CA WR, Giampietro PF, Wesbrook SD, Caldwell MD. Marshfield Clinic Personalized Medicine Research Project (PMRP): design, methods and recruitment for a large population-based biobank. *Personalized Medicine.* 2005; 2:49–79.
7. Lindstrom J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care.* 2003; 26:725–31. [PubMed: 12610029]
8. Kho AN, Hayes MG, Rasmussen-Torvik L, Pacheco JA, Thompson WK, Armstrong LL, et al. Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study. *J Am Med Inform Assoc.* 2012; 19:212–8. [PubMed: 22101970]
9. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. *SIGKDD Explorations.* 2009; 11:10–8.
10. Nadeau C, Bengio Y. Inference for the generalization error. *Mach Learn.* 2003; 52:239–81.
11. Heckerman D, Geiger D, Chickering DM. Learning Bayesian networks: the combination of knowledge and statistical data. *Mach Learn.* 1995; 20:197–243.
12. LeCessie S, Van Houwelingen JC. Ridge estimators in logistic regression. *Appl Stat.* 1992; 41:191–201.
13. Wilson PW, 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; 167:1068–74. [PubMed: 17533210]
14. Walford GA, Porneala BC, Dauriz M, Vassy JL, Cheng S, Rhee EP, et al. Metabolite traits and genetic risk provide complementary information for the prediction of future type 2 diabetes. *Diabetes Care.* 2014; 37:2508–14. [PubMed: 24947790]
15. Klein MS, Shearer J. Metabolomics and Type 2 Diabetes: Translating Basic Research into Clinical Application. *J Diabetes Res.* 2016; 2016:3898502. [PubMed: 26636104]
16. Wurtz P, Tiainen M, Makinen VP, Kangas AJ, Soininen P, Saltevo J, et al. Circulating metabolite predictors of glycemia in middle-aged men and women. *Diabetes Care.* 2012; 35:1749–56. [PubMed: 22563043]

17. Gall WE, Beebe K, Lawton KA, Adam KP, Mitchell MW, Nakhle PJ, et al. alpha-hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a nondiabetic population. *PLoS One*. 2010; 5:e10883. [PubMed: 20526369]
18. Sone H, Shimano H, Ebinuma H, Takahashi A, Yano Y, Iida KT, et al. Physiological changes in circulating mannose levels in normal, glucose-intolerant, and diabetic subjects. *Metabolism*. 2003; 52:1019–27. [PubMed: 12898467]
19. Dungan KM. 1,5-anhydroglucitol (GlycoMark) as a marker of short-term glycemic control and glycemic excursions. *Expert Rev Mol Diagn*. 2008; 8:9–19. [PubMed: 18088226]
20. Suhre K, Meisinger C, Doring A, Altmaier E, Belcredi P, Gieger C, et al. Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. *PLoS One*. 2010; 5:e13953. [PubMed: 21085649]
21. Ferrannini E, Natali A, Camastra S, Nannipieri M, Mari A, Adam KP, et al. Early metabolic markers of the development of dysglycemia and type 2 diabetes and their physiological significance. *Diabetes*. 2013; 62:1730–7. [PubMed: 23160532]
22. Nikiforova VJ, Giesbertz P, Wiemer J, Bethan B, Looser R, Liebenberg V, et al. Glyoxylate, a new marker metabolite of type 2 diabetes. *J Diabetes Res*. 2014; 2014:685204. [PubMed: 25525609]
23. Lillioja S, Mott DM, Spraul M, Ferraro R, Foley JE, Ravussin E, et al. Insulin resistance and insulin secretory dysfunction as precursors of non-insulin-dependent diabetes mellitus. Prospective studies of Pima Indians. *N Engl J Med*. 1993; 329:1988–92. [PubMed: 8247074]
24. Ludvigsson J, Carlsson A, Forsander G, Ivarsson S, Kockum I, Lernmark A, et al. C-peptide in the classification of diabetes in children and adolescents. *Pediatr Diabetes*. 2012; 13:45–50. [PubMed: 21910810]
25. Baron AD, Schaeffer L, Shragg P, Kolterman OG. Role of hyperglucagonemia in maintenance of increased rates of hepatic glucose output in type II diabetics. *Diabetes*. 1987; 36:274–83. [PubMed: 2879757]
26. Haeusler RA, Astiarraga B, Camastra S, Accili D, Ferrannini E. Human insulin resistance is associated with increased plasma levels of 12alpha-hydroxylated bile acids. *Diabetes*. 2013; 62:4184–91. [PubMed: 23884887]
27. Forouhi NG, Koulman A, Sharp SJ, Imamura F, Kroger J, Schulze MB, et al. Differences in the prospective association between individual plasma phospholipid saturated fatty acids and incident type 2 diabetes: the EPIC-InterAct case-cohort study. *Lancet Diabetes Endocrinol*. 2014; 2:810–8. [PubMed: 25107467]
28. Spranger J, Kroke A, Mohlig M, Bergmann MM, Ristow M, Boeing H, et al. Adiponectin and protection against type 2 diabetes mellitus. *Lancet*. 2003; 361:226–8. [PubMed: 12547549]
29. Van Woudenberg GJ, Kuijsten A, Van der Kallen CJ, Van Greevenbroek MM, Stehouwer CD, Blaak EE, et al. Comparison of fatty acid proportions in serum cholesteryl esters among people with different glucose tolerance status: the CoDAM study. *Nutr Metab Cardiovasc Dis*. 2012; 22:133–40. [PubMed: 20888206]
30. Rotter V, Nagaev I, Smith U. Interleukin-6 (IL-6) induces insulin resistance in 3T3-L1 adipocytes and is, like IL-8 and tumor necrosis factor-alpha, overexpressed in human fat cells from insulin-resistant subjects. *J Biol Chem*. 2003; 278:45777–84. [PubMed: 12952969]
31. Halaas JL, Gajiwala KS, Maffei M, Cohen SL, Chait BT, Rabinowitz D, et al. Weight-reducing effects of the plasma protein encoded by the obese gene. *Science*. 1995; 269:543–6. [PubMed: 7624777]
32. Considine RV, Sinha MK, Heiman ML, Kriauciunas A, Stephens TW, Nyce MR, et al. Serum immunoreactive-leptin concentrations in normal-weight and obese humans. *N Engl J Med*. 1996; 334:292–5. [PubMed: 8532024]
33. Hukshorn CJ, Saris WH, Westerterp-Plantenga MS, Farid AR, Smith FJ, Campfield LA. Weekly subcutaneous pegylated recombinant native human leptin (PEG-OB) administration in obese men. *J Clin Endocrinol Metab*. 2000; 85:4003–9. [PubMed: 11095423]
34. Rossetti L, Hawkins M, Chen W, Gindi J, Barzilai N. In vivo glucosamine infusion induces insulin resistance in normoglycemic but not in hyperglycemic conscious rats. *J Clin Invest*. 1995; 96:132–40. [PubMed: 7615783]

35. Simon RR, Marks V, Leeds AR, Anderson JW. A comprehensive review of oral glucosamine use and effects on glucose metabolism in normal and diabetic individuals. *Diabetes Metab Res Rev*. 2011; 27:14–27. [PubMed: 21218504]
36. Nestel PJ, Straznicky N, Mellett NA, Wong G, De Souza DP, Tull DL, et al. Specific plasma lipid classes and phospholipid fatty acids indicative of dairy food consumption associate with insulin sensitivity. *Am J Clin Nutr*. 2014; 99:46–53. [PubMed: 24153346]
37. 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; 28:412–9. [PubMed: 3899825]
38. Lemaitre RN, Fretts AM, Sitlani CM, Biggs ML, Mukamal K, King IB, et al. Plasma phospholipid very-long-chain saturated fatty acids and incident diabetes in older adults: the Cardiovascular Health Study. *Am J Clin Nutr*. 2015; 101:1047–54. [PubMed: 25787996]
39. Weyer C, Funahashi T, Tanaka S, Hotta K, Matsuzawa Y, Pratley RE, et al. Hypoadiponectinemia in obesity and type 2 diabetes: close association with insulin resistance and hyperinsulinemia. *J Clin Endocrinol Metab*. 2001; 86:1930–5. [PubMed: 11344187]
40. Wannamethee SG, Tchernova J, Whincup P, Lowe GD, Rumley A, Brown K, et al. Associations of adiponectin with metabolic and vascular risk parameters in the British Regional Heart Study reveal stronger links to insulin resistance-related than to coronary heart disease risk-related parameters. *Int J Obes (Lond)*. 2007; 31:1089–98. [PubMed: 17264850]
41. Menni C, Fauman E, Erte I, Perry JR, Kastenmuller G, Shin SY, et al. Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach. *Diabetes*. 2013; 62:4270–6. [PubMed: 23884885]

Fig. 1. Receiver operating characteristic curves generated using (A) a Bayesian network classifier and (B) ridge logistic regression to distinguish incident type 2 diabetes cases ($n = 56$) from non-diabetic controls ($n = 445$) in the validation group. Curves were generated for each of three classifiers: glucose only (Glucose), metabolites in panel 7 only (Metabolites), and the combination of metabolites in panel 7 and six metabolic disease-related proteins (Metabolites + Proteins). Values for the area under the receiver

operating characteristic curves (AUCs) are color-coded according to the classifier. No statistically significant differences were observed for pairwise comparisons of the AUCs.

Table 1

Characteristics of study participants.

Study group	n (percent female)	Age (years) ^a	Body mass index (kg/m ²) ^a	Fasting plasma glucose (mean pool- normalized ratio) ^{a,b}
Biomarker discovery ^c				
Type 2 diabetes	61 (21)	55.2 (53.5, 56.9)	30.9 (29.9, 31.9)	1.14 (1.10, 1.17)
Controls	78 (28)	54.1 (52.5, 55.8)	28.6 (27.8, 29.4)	0.92 (0.90, 0.94)
Biomarker validation ^d				
Type 2 diabetes	56 (38)	60.0 (56.7, 63.2)	34.2 (32.4, 36.0)	1.38 (1.28, 1.48)
Controls	445 (49)	55.5 (54.3, 56.7)	33.1 (32.6, 33.7)	0.98 (0.97, 1.00)

^aMean (95% confidence interval) measured at study enrollment.^bRelative abundance measured by metabolite profiling.^cSubjects from the Bavarian Red Cross Blood Transfusion Center Study.^dSubjects from the Personalized Medicine Research Project.

Table 2

Comparison between AUCs for metabolite panels and AUC for plasma glucose alone.

Metabolite panel number	Classification model	AUC for metabolite panel ^a	Difference between AUC for metabolite panel and AUC for glucose alone	p value ^b
1 – 4	Random forest	< 0.866	–0.015 to 0.022	$> 1.60 \times 10^{-1}$
5	Elastic net	0.897	0.030	1.80×10^{-4}
6	Elastic net	0.890	0.024	8.81×10^{-3}
7	Elastic net	0.899	0.033	2.08×10^{-5}

Abbreviations: AUC, area under the receiver operating characteristic curve.

^aThe AUC analysis was performed using the biomarker discovery group (61 type 2 diabetes cases and 78 non-diabetic controls). The analysis compared the performance of metabolite panels with that of plasma glucose alone to distinguish type 2 diabetes cases from non-diabetic controls. AUC for plasma glucose alone = 0.866.

^bCorrected re-sampled *t*-test used to compare AUCs of metabolite panels with AUC of plasma glucose alone.

Table 3
ANOVA comparing concentrations of metabolites in panel 7 between type 2 diabetes cases and controls.

Metabolite	Metabolite ontological class	Biomarker discovery group ^a		Biomarker validation group ^b	
		Ratio ^{c,d}	p value ^d	Ratio ^{c,d}	p value ^d
Glucose	Monosaccharides	1.23	2.41×10^{-20}	1.36	2.24×10^{-28}
Mannose	Monosaccharides	1.31	2.37×10^{-8}	1.33	6.40×10^{-17}
Glucosamine	Aminosugars	1.57	1.01×10^{-7}	1.45	1.16×10^{-7}
2-Hydroxybutyrate	Energy metabolism – carboxylic acid	1.22	1.59×10^{-3}	1.15	1.28×10^{-2}
Glyoxylate	Carboxylic acid	1.24	2.13×10^{-2}	1.31	4.16×10^{-6}
Cholesteryl ester (C18:1)	Cholesteryl esters	0.88	3.02×10^{-3}	0.77	3.09×10^{-5}
Glycochenodeoxycholic acid	Bile acids	1.64	1.17×10^{-2}	1.28	1.18×10^{-1}
Taurochenodeoxycholic acid	Bile acids	1.76	3.21×10^{-3}	1.78	3.37×10^{-4}
Lysophosphatidylcholine (C18:2)	Lysophosphatidylcholine	0.92	4.45×10^{-3}	0.96	1.42×10^{-1}
Lignoceric acid (C24:0)	Fatty acids, saturated	0.95	1.83×10^{-1}	0.93	7.29×10^{-2}

Abbreviations: ANOVA, analysis of variance.

^a 61 type 2 diabetes cases and 78 non-diabetic controls from the Bavarian Red Cross Blood Transfusion Center study.

^b 56 type 2 diabetes cases and 445 non-diabetic controls from the Personalized Medicine Research Project.

^c Ratio comparing plasma metabolite pool-normalized ratios between type 2 diabetes cases and non-diabetic controls. The pool-normalized ratio is the relative abundance measured by metabolite profiling.

^d Ratios and p values from ANOVA models adjusted for age at study enrollment, sex, body mass index, interaction of sex and age, and interaction of sex and body mass index.

Table 4

Comparison of plasma concentrations of metabolic disease-related proteins between type 2 diabetes cases and controls in the biomarker validation group.

Protein	Cases (n = 56) ^a	Controls (n = 445) ^a	Ratio ^{b,c}	p value ^c
C-peptide (ng/mL)	3.77 (3.35, 4.20)	2.30 (2.21, 2.40)	1.54	1.18×10^{-13}
Insulin (ng/mL)	0.95 (0.76, 1.15)	0.49 (0.46, 0.53)	1.78	5.21×10^{-12}
Adiponectin (ng/mL)	11.90 (9.92, 13.88)	14.13 (13.38, 14.87)	0.81	5.63×10^{-4}
Interleukin-6 (pg/mL)	1.61 (1.23, 2.00)	1.68 (0.88, 2.48)	1.36	1.25×10^{-3}
Glucagon (pg/mL)	90.77 (71.25, 110.30)	68.33 (64.64, 72.01)	1.27	3.74×10^{-3}
Leptin (ng/mL)	28.45 (20.95, 35.95)	26.89 (24.20, 29.59)	1.15	2.45×10^{-1}

^aValues are mean (95% confidence interval).

^bRatio comparing plasma protein concentrations between type 2 diabetes cases and non-diabetic controls.

^cRatios and p values from analysis of variance (ANOVA) models adjusted for age at study enrollment, sex, body mass index, interaction of sex and age, and interaction of sex and body mass index.

Table 5

AUC for classifying patients with type 2 diabetes and controls in the biomarker validation group.

Features used for classification	AUC ^a	
	Bayesian network	Ridge logistic regression
Glucose alone	0.88 ± 0.09	0.88 ± 0.08
Metabolites in panel 7	0.89 ± 0.09	0.90 ± 0.06
Metabolites in panel 7 + proteins	0.91 ± 0.06	0.89 ± 0.07
Metabolites in panel 7 (without glucose)	0.80 ± 0.10	0.86 ± 0.07
Metabolites in panel 7 (without glucose) + proteins	0.87 ± 0.07	0.87 ± 0.07
Model of Wilson et al. (clinical characteristics) ^b	0.87 ± 0.07	0.86 ± 0.08
Model of Wilson et al. + metabolites in panel 7	0.90 ± 0.07	0.89 ± 0.07
Model of Wilson et al. + metabolites in panel 7 + proteins	0.92 ± 0.06 ^c	0.89 ± 0.07
Model of Walford et al. (selected amino acids + selected lipids + genetic risk score) ^d	0.63 ± 0.10	0.71 ± 0.13
Model of Walford et al. + metabolites in panel 7	0.87 ± 0.09 ^e	0.91 ± 0.06 ^f
Model of Walford et al. + metabolites in panel 7 + proteins	0.90 ± 0.06 ^g	0.91 ± 0.06 ^h
Models of Wilson et al. and Walford et al. + metabolites in panel 7	0.89 ± 0.07	0.89 ± 0.07
Models of Wilson et al. and Walford et al. + metabolites in panel 7 + proteins	0.91 ± 0.06	0.89 ± 0.07

Abbreviation: AUC, area under the receiver operating characteristic curve.

^aValues are average AUC ± standard deviation from 10-fold cross validation.

^bModel of Wilson et al. [13] included age at study enrollment, sex, body mass index, family history of type 2 diabetes in a first-degree relative, systolic blood pressure at or near study enrollment, fasting plasma glucose, plasma high density lipoprotein-cholesterol, and plasma triglycerides. We used family history of type 2 diabetes in a first-degree relative as a substitute for parental history of type 2 diabetes.

^c $p = 3.96 \times 10^{-2}$ for comparison with model of Wilson et al.

^dModel reported by Walford et al. [14] included age at study enrollment, sex, genetic risk score based on 62 single nucleotide polymorphisms, and fasting plasma concentrations of isoleucine, tyrosine, phenylalanine, lysophosphatidylcholine C18:2, phosphatidylcholine C38:6, triacylglycerol C44:1, triacylglycerol C48:0, triacylglycerol C52:1, and triacylglycerol C56:9. In our study population, values were not available for the last five metabolites; therefore, we did not include these five metabolites in the model.

^e $p = 4.47 \times 10^{-10}$ for comparison with model of Walford et al.

^f $p = 1.86 \times 10^{-5}$ for comparison with model of Walford et al.

^g $p = 1.18 \times 10^{-12}$ for comparison with model of Walford et al.

^h $p = 1.03 \times 10^{-5}$ for comparison with model of Walford et al.