Identification of Serum Metabolites Associated With Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach

Anna Floegel,1 Norbert Stefan,2 Zhonghao Yu,3 Kristin Mühlenbruch,4 Dagmar Drogan,1 Hans-Georg Joost,5 Andreas Fritsche,2 Hans-Ulrich Häring,2 Martin Hrabě de Angelis,6 Annette Peters,7 Michael Roden,8,9 Cornelia Prehn,6 Rui Wang-Sattler,3 Thomas Illig,3,10 Matthias B. Schulze,4 Jerzy Adamski,6 Heiner Boeing,1 and Tobias Pischon1,11

Metabolomic discovery of biomarkers of type 2 diabetes (T2D) risk may reveal etiological pathways and help to identify individuals at risk for disease. We prospectively investigated the association between serum metabolites measured by targeted metabolomics and risk of T2D in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam (27,548 adults) among all incident cases of T2D (n = 800, mean follow-up 7 years) and a randomly drawn subcohort (n = 2,282). Flow injection analysis tandem mass spectrometry was used to quantify 163 metabolites, including acylcarnitines, amino acids, hexose, and phospholipids, in baseline serum samples. Serum hexose; phenylalanine; and diacyl-phosphatidylcholines C32:1, C36:1, C38:3, and C40:5 were independently associated with increased risk of T2D and serum glycine; sphingomyelin C16:1; acyl-alkyl-phosphatidylcholines C34:3, C40:6, C42:5, C44:4, and C44:5; and lysosphosphatidylcholine C18:2 with decreased risk. Variance of the metabolites was largely explained by two metabolite factors with opposing risk associations (factor 1 relative risk in extreme quintiles 0.31 [95% CI 0.21–0.44], factor 2 3.82 [2.64–5.52]). The metabolites significantly improved T2D prediction compared with established risk factors.

Type 2 diabetes (T2D) is characterized by impaired insulin sensitivity of several tissues and inadequate insulin secretion from β-cells (1). A detailed understanding of the pathophysiology of T2D is a prerequisite for the development of preventive strategies. In particular, the identification of early metabolic alterations is promising in the study of etiological pathways and may further help to identify high-risk individuals. A number of biomarkers have been proposed as indicators for the estimation of T2D risk, such as fasting plasma glucose and glycated hemoglobin A1c (HbA1c) (2), triglycerides (3), HDL cholesterol (4), inflammatory markers (5), adiponectin (6), liver enzymes (7), and fetuin-A (8). However, most biomarkers fail to grasp the complexity of T2D etiology (3). Design and advancement of high-throughput analytical techniques determined the emergence of metabolomics, which is the simultaneous study of numerous low-molecular weight compounds, namely metabolites. Metabolites represent intermediates and end products of metabolic pathways that reflect more rapidly physiological dysfunctions than current biomarkers and, thus, may mirror earlier stages of T2D (9). Cross-sectional studies have linked alterations in metabolic profiles with obesity (10), glucose tolerance (11), and prevalent diabetes (12–14). The most prominent metabolic shifts involved blood acylcarnitines and branched-chain amino acids (BCAAs). Recently, observations from a prospective study found that a set of five amino acids was predictive for T2D, representing pioneering work in the emerging field of systems epidemiology (15).

In the current study, we investigated whether a targeted metabolomic approach involving a broader spectrum of metabolites and a larger number of study participants may help to identify metabolites associated with the risk of T2D and the mechanisms involved. Therefore, we profiled 163 serum metabolites in originally healthy individuals who were consecutively followed up for incident T2D in two large-scale prospective cohort studies in Germany and studied cross-sectional relationships of the identified metabolites with insulin sensitivity and secretion in precisely phenotyped participants. In addition, we evaluated the usefulness of the metabolites for T2D risk prediction compared with the German Diabetes Risk Score (DRS) (16) and established biomarkers.
RESEARCH DESIGN AND METHODS
European Prospective Investigation into Cancer and Nutrition—Potsdam study. The European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam is part of the ongoing multicenter EPIC study and comprises 27,548 participants from the general population in the area of Potsdam in eastern Germany who were mainly 35–65 years of age at time of recruitment between 1994 and 1998 (17). At baseline, participants underwent an examination by qualified staff, including medical history, blood pressure measurement, and anthropometry (18). Participants also completed sociodemographic and lifestyle questionnaires and a validated food frequency questionnaire. In addition, 30 mL blood were drawn (random sampling) and immediately processed (19). Only participants with morning appointments were asked to fast overnight. Blood was fractionated into serum, plasma, buffy coat, and erythrocytes; aliquots of 0.5 mL each were stored in tanks of liquid nitrogen at −196°C until analysis. Besides metabolomic profiling, other biomarkers have been measured in baseline blood samples as described previously (7,20,21).

Every 2–3 years, follow-up questionnaires were sent to participants to identify incident cases of T2D, with response rates of ~90% (22,23). Once a participant was identified as a potential case, disease status was further verified with medical records, including the correct diagnosis (International Classification of Diseases, 10th revision, E11, non–insulin-dependent diabetes), the date of the diagnosis, and the means of diagnosis confirmation. This verification was achieved by sending a standard inquiry form to the treating physician. Consent was obtained from all study participants a priori, and the study was approved by the ethics committee of the Medical Society of the State of Brandenburg.

We constructed a case-cohort study within EPIC-Potsdam, including all incident cases of T2D of the full cohort identified up to 31 August 2005 (n = 849, mean age = 63 years), and a random sample of n = 2,500 participants randomly drawn from the EPIC-Potsdam study population. Because the subcohort was representative of the full cohort, it included 2,415 noncases and 85 cases of incident T2D (i.e., internal cases). The remaining 764 cases did not belong to the subcohort (external cases). By randomly selecting the subcohort and using the appropriate statistic for this study design, the biomarkers only needed to be measured in the case-cohort sample; however, the results are expected to be generalizable to the full cohort (7). The case-cohort design was previously chosen based on its advantages, including a reduced chance of selection bias for the control group (24).

For the present analysis, we further excluded participants with prevalent T2D at baseline (n = 110), with missing or nonverified data on incident or prevalent T2D (n = 13), with missing blood samples or biomarker measurements (n = 64), and with missing covariate information (n = 80). Thus, the analyzable sample comprised 2,282 individuals of the subcohort and 800 individuals with incident T2D.

Cooperative Health Research in the Region of Augsburg study. The Cooperative Health Research in the Region of Augsburg (KORA) study consists of population-based surveys and follow-up periods in the area of Augsburg in southern Germany. A total of 4,201 individuals between 25 and 74 years of age participated in the S4 survey between 1999 and 2001 (25). In the KORA cohort, blood was drawn into serum gel tubes after a fasting period of at least 8 h. Blood samples were rested for coagulation for 30 min at room temperature; serum was obtained by centrifugation at 2,750 g at 15°C for 10 min and stored in a freezer at −80°C. A total of 3,080 individuals took part in the F4 follow-up survey during the years 2006–2008 (26). The identification of incident T2D was based on an oral glucose tolerance test (OGTT) or a validated physician diagnosis (27). All KORA participants gave written informed consent, and the KORA study was approved by the ethics committee of the Bavarian medical association.

A subcohort of 876 S4 participants 55–74 years of age without T2D at baseline and with metabolomics data available was included in the current study. Of them, 91 developed incident T2D during the 7-year follow-up.

Tübingen Family study for T2D. The Tübingen Family (TuF) study is in an ongoing investigation of the pathophysiology of T2D in southern Germany (28). Individuals meeting at least one of the following criteria were included in the study: a family history of T2D, a BMI ≥27 kg/m², and previous impaired glucose tolerance or gestational diabetes mellitus. They were considered healthy according to a physical examination and routine laboratory tests. Written informed consent was obtained from all participants, and the medical ethics committee of the University of Tübingen approved the protocol. All individuals underwent a 75-g OGTT. Venous plasma samples were drawn at 0, 30, 60, and 120 min for plasma glucose, insulin, C-peptide, and metabolic analyses (minute 0). Insulin secretion was calculated from the OGTT (29). The plasma glucose and C-peptide areas under the curve (AUCs) during the OGTT were calculated by applying the trapezoidal method. Insulin secretion was calculated from AUC_C-peptide/AUC_glucose. The present analysis included 76 Caucasians from the TuF study who had measurements of insulin sensitivity and insulin secretion as well as metabolomics data available.

Serum metabolite concentrations. Serum concentrations of metabolites were determined with the Absolute/DEQ p150 and p180 Kits (Biocrates Life Sciences AG, Innsbruck, Austria) using the flow injection analysis tandem mass spectrometry (FIA-MS/MS) technique (30). The metabolomic method specifically quantified 163 metabolites, including 41 acylcarnitines (Cxc), 14 amino acids, 1 hexose (sum of six-carbon monosaccharides without distinction of isomers), 92 glycerophospholipids (lyso-, diacyl-, and acyl-alkyl-phosphatidylcholines), and 15 sphingomyelins. To ensure valid measurements, metabolites below the limit of detection (n = 30) and those with very high analytical variance (n = 6) in our samples were excluded, leaving 127 metabolites for the present analysis.

Metabolomic measurements were performed in the Genome Analysis Center at the Helmholtz Zentrum München. Sample preparation was done according to the manufacturer’s protocol (Biocrates user’s manual UM-P150) and has been described previously (30). In brief, after centrifugation, 10 μL serum were inserted into a filter on a 96-well sandwich plate, which already contained stable isotope-labeled internal standards. Amino acids were derivatized with 5% phenylisothiocyanate reagent. Metabolites and internal standards were extracted with 5 mmol/L ammonium acetate in methanol. The solution was then centrifuged through a filter membrane and diluted with mass spectrometry running solvent. Final extracts were analyzed by FIA-MS/MS, and metabolites were quantified in μmol/L by appropriate internal standards. The method has been validated, and analytical specifications were provided in the Biocrates User’s Manual (UM-P150). The metabolite selection was based on the robustness of their measurements. The uncertainty of the measurements was <10% for most of the metabolites. Regarding accuracy, all included metabolites were found in the range of 80–115% of their theoretical values. The median analytical variance of EPIC-Potsdam samples was a 7.3% within-plate coefficient of variation and a 11.3% between-plates coefficient of variation (31). To account for run-order effects, serum samples were randomly analyzed together, regardless of the case status. We have shown previously that most of the metabolites had moderate to high intraclass correlation coefficients measured in participants over a 4-month period, indicating reasonable reliability of the measurements (31).

Statistical analysis
Step 1: identification of metabolites associated with T2D risk in EPIC-Potsdam. Cox proportional hazards regression with weighting as suggested by Wilkerson (32) was performed to robustly select covariates and variance explainers. For the case-cohort design were used to calculate multivariable-adjusted hazard ratios as a measure of relative risk (RR) and 95% CI, with age as the underlying time scale from recruitment to study exit (T2D diagnosis or censoring) of each participant. We considered z score–standardized metabolite concentrations (mean 0 [SD 1]) as the exposure variable and calculated a multivariable-adjusted model to select metabolites associated with T2D risk. This model was adjusted for age, sex, alcohol intake from beverages (nonconsumers: women ≤0–6, 6–12, and >12 g/day; and men >0–12, 12–24, and >24 g/day), smoking (never, former, current ≤20 cigarettes/day, current >20 cigarettes/day), cycling and sports (h/week), education (no degree/vocational training, trade/technical school, university degree), coffee intake (cups/day), red meat intake (g/day), whole-grain bread intake (g/day), prevalent hypertension (yes/no), waist circumference (cm), and BMI. Because the metabolomic approach is exploratory, the P values from Cox regression were corrected for multiple testing (n = 127) using the Bonferroni-Holm procedure (33), and a corrected P < 0.05 (two-sided testing) was considered significant to select metabolites.

We next calculated a model that included the covariates and all the identified metabolites and used stepwise Cox regression to select the independent predictors. To also account for the intercorrelation of some of the metabolites, we conducted a principal component analysis (PCA). In brief, the PCA aggregates the individual metabolites based on their degree of correlation with one another to a smaller number of metabolite factors (principal components). These metabolite factors are extracted in a way that explains the major fraction of the variance of individual metabolites. We included all metabolites associated with T2D risk in the PCA, based the PCA on the correlation matrix and used a principal orthogonal rotation procedure with the varimax method. We retained two metabolite factors according to the scree test and because they accounted for most of the observed variance. Thus, the proportion of explained variance of factor 1 and factor 2 were 34.2 and 16.1%, respectively. To investigate the association between metabolite factors and T2D risk, we estimated RR and 95% CI across quintiles of metabolite factors, particularly choosing quintiles of metabolite factor 2 that lie above the interpretation. We also investigated a possible effect modification of sex or fasting status on the association between metabolite factors and T2D risk by including multiplicative interaction terms into the models. We then repeated the PCA to include only fasting blood samples in order to evaluate whether the metabolite factors were different from those obtained from random blood samples. Finally, we calculated hazard ratios of T2D during different periods of follow-up and tested whether they were different with a test of heterogeneity (34).
Step 2: additional analyses, risk prediction, and replication in KORA.

For significant metabolites found in step 1, we calculated multivariable-adjusted models with additional adjustment for blood glucose, HbA1c, HDL cholesterol, and triglycerides. We also adjusted for the amino acids phenylalanine, tyrosine, and isoleucine, which have been found to be associated with T2D risk (35). Using data of the EPIC-Potsdam subcohort, we calculated Spearman partial correlation coefficients between identified metabolites and established T2D biomarkers. Data of the TuF study were used to calculate Spearman partial correlation coefficients between identified metabolites and measures of insulin sensitivity and secretion. We calculated measures of discrimination and calibration in different multivariable-adjusted logistic regression models using the DRS (16) as the reference model and adding established T2D biomarkers and the identified metabolites. Receiver operating characteristic (ROC) AUCs were compared using the DeLong test (36).

The results were replicated in the prospective KORA study, and metabolite factors were recalculated in KORA using the linear factor equations retrieved from the PCA in the EPIC-Potsdam sample. The risk estimates from EPIC-Potsdam and KORA were combined using a meta-analytical approach (37). Power calculation suggested a detectable RR per SD of 1.26 (38). Additionally, multivariable-adjusted RRs of T2D were calculated for the amino acids that were recently identified by Wang et al (35) and that were also measured in the EPIC-Potsdam study. The statistical analyses were conducted with SAS version 9.2 (SAS Institute, Inc, Cary, NC) and in the R statistical environment (www.r-project.org).

**RESULTS**

Baseline characteristics of the EPIC-Potsdam sample are presented in Table 1. Of all the metabolites measured using a targeted metabolomic approach, 34 were significantly associated with T2D risk in the EPIC-Potsdam study after correction for multiple testing (Supplementary Table 1). These relations were also independent of relevant dietary and lifestyle factors as well as anthropometry and hypertension. Among these 34 metabolites, 14 were identified to be significantly associated with T2D risk independent of the others (Table 2). Specifically, hexose, phenylalanine, and diacyl-phosphatidylcholines C32:1, C36:1, C38:3, and C40:5 were significantly positively associated with T2D risk, whereas glycine, sphingomyelin C16:1, lysophosphatidylcholine C18:2, and acyl-alkyl-phosphatidylcholines C34:3, C40:6, C42:5, C44:4, and C44:5 were significantly inversely related to T2D risk. Using PCA, we identified two metabolite factors that included multiple metabolites and explained most of their variation (Fig. 1). These metabolite factors showed significant and opposing associations with T2D risk. When comparing extreme quintiles, metabolite factor 1, which mainly contains acyl-alkyl-phosphatidylcholines, sphingomyelins, and lysophosphatidylcholines, was associated with a significant 69% reduced risk of T2D (RR 0.31 [95% CI 0.21–0.44]) (Table 3), whereas metabolite factor 2, consistent of diacyl-phosphatidylcholines, BCAA and aromatic amino acids, propionylcarnitine, and hexose, was associated with a significant 3.82-fold increased risk of T2D (3.82 [2.64–5.52]). When we restricted the PCA to fasting samples (n = 429), very similar metabolite factors could be generated with only minor differences in factor loadings (Supplementary Table 2). We also estimated the joint effects of both metabolite factors by summing them (factor 1 received a negative sign because it was inversely associated with T2D risk) and calculating RR of T2D across quintiles of combined factors. The RR (95% CI) of T2D from quintile 1 to quintile 5 of summed metabolite factors was as follows: 1.0, 1.47 (0.96–2.24), 2.29 (1.54–3.38), 2.67 (1.79–4.0), and 6.69 (4.50–9.96) (P for trend < 0.0001).

Adjustment for established T2D biomarkers only marginally affected the magnitude of risk association for most

### Table 1

Baseline characteristics of the EPIC-Potsdam case-cohort sample (1994–1998)

| Age (years)‡ | Subcohort (n = 2,282) | Incident T2D cases (n = 800) | P for difference† |
|--------------|-----------------------|-----------------------------|-------------------|
| 49.5 (8.9)   | 54.7 (7.3)            | 0.0001                      |
| Women (%)‡   | 62.0                  | 42.2                        | 0.0001            |
| BMI (kg/m²)  | 26.1 (0.69)           | 30.1 (0.15)                 | 0.0001            |
| Waist circumference, men (cm)§ | 93.7 (0.34) | 103.6 (0.46) | 0.0001 |
| Waist circumference, women (cm)§ | 80.6 (0.30) | 93.4 (0.62) | 0.0001 |
| Prevalent hypertension | 49.5 | 70.8 | 0.0001 |
| Education | No degree/vocational training | 37.1 | 45.6 | 0.0001 |
| | Trade/technical school | 24.0 | 25.4 | 0.412 |
| | University degree | 39.0 | 29.0 | 0.0001 |
| Smoking status | Never | 46.9 | 36.2 | 0.0001 |
| | Former | 33.0 | 42.3 | 0.0001 |
| Current | 20.1 | 21.5 | 0.472 |
| Among smokers, cigarettes/day | 12.6 (0.43) | 16.0 (0.74) | 0.0001 |
| Physical activity (h/week)¶ | 2.8 (0.07) | 2.2 (0.13) | 0.0001 |
| Alcohol intake from beverages (g/day) | 14.8 (0.41) | 14.5 (0.71) | 0.761 |
| Coffee consumption (cups/day) | 2.8 (0.04) | 2.7 (0.08) | 0.148 |
| Whole-grain bread intake (g/day) | 45.9 (1.11) | 38.2 (1.91) | 0.0003 |
| Red meat intake (g/day) | 43.3 (0.61) | 48.9 (1.06) | 0.0001 |
| Biomarkers | Glucose (mg/dL) | 88.1 (0.53) | 107.0 (0.92) | 0.0001 |
| | HbA1c (%) | 5.42 (0.01) | 6.30 (0.03) | 0.0001 |
| | HDL cholesterol (mg/dL) | 47.5 (0.25) | 40.9 (0.43) | 0.0001 |
| | Triglycerides (mg/dL) | 114.8 (2.12) | 177.2 (3.65) | 0.0001 |

Data are age- and sex-adjusted mean (SE) for continuous variables or % for categorical variables. †P for difference comparing incident cases of T2D with noncases of the subcohort. ‡Unadjusted mean (SD) or % §Age-adjusted mean (SE). ¶Average of cycling and sports during summer and winter seasons. ††Data were only available for n = 2,900.
TABLE 2
Mean serum concentrations of identified metabolites and their association with risk of T2D in EPIC-Potsdam Subcohort (n = 2,282) (µmol/L)

| Subcohort | Cases (n = 800) (µmol/L) | RR per SD (95% CI) |
|-----------|--------------------------|--------------------|
| Hexose    | 4,698 (984) 5,783 (2,167) | Basic model† + Glucose‡ + HbA1c‡ + HDL chol‡ + Triglycerides‡ + Glucose + HbA1c + HDL chol + Triglycerides |
| Phenylnaline | 56.4 (12.0) 61.0 (12.9) | 1.35 (1.22–1.49) 1.24 (1.12–1.37) 1.37 (1.21–1.52) 1.34 (1.21–1.49) 1.30 (1.16–1.44) 1.25 (1.12–1.39) |
| Glycine    | 256 (77.2) 227 (59.6) | 0.73 (0.64–0.83) 0.81 (0.71–0.92) 0.81 (0.71–0.93) 0.74 (0.65–0.84) 0.74 (0.65–0.84) 0.89 (0.78–1.01) |
| SM C16:1  | 17.6 (3.86) 17.4 (4.02) | 0.80 (0.72–0.88) 0.82 (0.73–0.92) 0.74 (0.67–0.83) 0.87 (0.78–0.97) 0.79 (0.71–0.88) 0.91 (0.80–1.03) |
| PC ae C34:3 | 8.51 (2.48) 7.01 (2.02) | 0.64 (0.56–0.72) 0.70 (0.61–0.80) 0.67 (0.58–0.77) 0.72 (0.63–0.82) 0.66 (0.58–0.74) 0.92 (0.78–1.07) |
| PC ae C40:6 | 5.52 (1.43) 5.03 (1.34) | 0.77 (0.69–0.86) 0.81 (0.72–0.92) 0.83 (0.74–0.94) 0.82 (0.73–0.92) 0.76 (0.67–0.85) 0.94 (0.82–1.08) |
| PC ae C42:5 | 2.35 (0.54) 2.06 (0.46) | 0.71 (0.63–0.79) 0.76 (0.67–0.86) 0.82 (0.72–0.92) 0.76 (0.68–0.86) 0.73 (0.64–0.82) 0.96 (0.83–1.12) |
| PC ae C44:4 | 0.38 (0.10) 0.33 (0.09) | 0.76 (0.67–0.85) 0.77 (0.68–0.87) 0.84 (0.75–0.96) 0.82 (0.72–0.93) 0.77 (0.68–0.87) 0.97 (0.84–1.12) |
| PC ae C44:5 | 1.77 (0.48) 1.53 (0.43) | 0.71 (0.63–0.80) 0.74 (0.65–0.84) 0.81 (0.71–0.91) 0.77 (0.68–0.87) 0.74 (0.65–0.83) 0.95 (0.82–1.09) |
| PC aa C32:1 | 16.9 (9.76) 20.4 (11.3) | 1.18 (1.08–1.29) 1.12 (1.02–1.24) 1.12 (1.02–1.23) 1.23 (1.11–1.36) 1.12 (1.02–1.22) 1.15 (1.03–1.29) |
| PC aa C36:1 | 57.3 (15.6) 62.1 (17.2) | 1.18 (1.08–1.30) 1.18 (1.07–1.30) 1.09 (0.98–1.21) 1.26 (1.14–1.39) 1.07 (0.96–1.19) 1.25 (1.10–1.41) |
| PC aa C38:3 | 55.4 (14.8) 65.3 (18.2) | 1.37 (1.24–1.51) 1.35 (1.22–1.50) 1.31 (1.18–1.45) 1.38 (1.25–1.52) 1.24 (1.11–1.38) 1.38 (1.22–1.55) |
| PC aa C40:5 | 11.4 (3.49) 12.8 (4.17) | 1.18 (1.08–1.29) 1.21 (1.10–1.33) 1.12 (1.02–1.24) 1.21 (1.10–1.32) 1.07 (0.96–1.18) 1.19 (1.06–1.32) |
| LysoPC a C18:2 | 35.2 (13.7) 29.6 (11.1) | 0.74 (0.66–0.84) 0.72 (0.64–0.81) 0.77 (0.68–0.88) 0.79 (0.70–0.90) 0.72 (0.63–0.81) 0.84 (0.73–0.96) |

Data are mean (SD) metabolite concentrations. a, acyl; aa, diacyl; ae, acyl-alkyl; chol, cholesterol; PC, phosphatidylcholine; SM, sphingomyelin. †Metabolites were selected by stepwise Cox proportional hazard regression modified by the Prentice method after standardizing metabolite concentrations to a mean of 0 (SD 1) and including all metabolites associated with T2D risk as well as relevant covariates. RRs for the basic model were then calculated in a continuous model adjusted for age, sex, alcohol intake from beverages (nonconsumers; women: 0–6, 6–12, and >12 g/day; men: 0–12, 12–24, and >24 g/day), smoking (never, former, current ≤20 cigarettes/day, current >20 cigarettes/day), physical activity (cycling and sports in h/week), education (low, medium, high), coffee intake (cups/day), red meat intake (g/day), whole-grain bread intake (g/day), prevalent hypertension (yes/no), BMI (kg/m²), and waist circumference (cm). ‡Adjustment according to the basic model with additional adjustment for specified biomarkers.
of the metabolites (Table 2). An exception was that the inverse association between acyl-alkyl-phosphatidylcholines and sphingomyelin C16:1 and T2D risk was attenuated and no longer significant after adjustment for blood glucose, HbA1c, HDL cholesterol, and triglycerides. The positive association between hexose and T2D risk was considerably weakened but remained significant after adjustment for plasma glucose. When we also adjusted for phenylalanine, tyrosine, and isoleucine, the associations for metabolite factor 1 were unchanged. The associations for metabolite factor 2, which included these three amino acids, were weakened but remained significant (data not shown).

We further observed that metabolites were linked to established T2D biomarkers (Supplementary Table 3). Specifically, metabolite factor 1, which was inversely associated with T2D risk, was negatively correlated to plasma glucose, HbA1c, and triglycerides and positively related to HDL cholesterol and adiponectin. Metabolite factor 2, which was positively associated with T2D risk, was positively correlated with triglycerides and liver enzymes. Data from the TuF study revealed that acyl-alkyl-phosphatidylcholines, lysophosphatidylcholine C18:2, and glycine were positively associated with insulin sensitivity, whereas hexose and diacyl-phosphatidylcholines were inversely related to insulin sensitivity (Table 4). Furthermore, phenylalanine was positively associated with insulin secretion, whereas hexose, sphingomyelin C16:1, and acyl-alkyl-phosphatidylcholines were inversely related to insulin secretion. The potential of identified metabolites to discriminate between T2D cases and noncases was comparable to that of the DRS (16) (ROC AUC 0.849 and 0.847, respectively, P for difference = 0.838) (Fig. 2). When the metabolites were added to established risk prediction models of T2D, discrimination was slightly but significantly improved up to a ROC AUC of 0.912, and these models were well calibrated (Fig. 2 and Supplementary Table 4). Replication in KORA revealed significant associations with T2D risk for metabolite factor 2 and hexose (Table 3 and Supplementary Table 5). In KORA, similar trends as in EPIC-Potsdam were seen for metabolite factor 1, acyl-alkyl-phosphatidylcholines, glycine, lysophosphatidylcholine C18:2, and sphingomyelin C16:1, with borderline significance. Although the risk estimates for diacyl-phosphatidylcholines were considerably lower in KORA than in EPIC-Potsdam, there was no significant heterogeneity between studies. We also calculated the RR of T2D for the BCAA and aromatic amino acids, which have recently been reported to be associated with T2D risk in the Framingham Offspring cohort and the Malmö Diet and Cancer study, to facilitate the comparison (35). In EPIC-Potsdam, isoleucine, valine, tyrosine, and phenylalanine were positively associated with T2D risk (RR per SD 1.30 [95% CI 1.17–1.43], 1.27 [1.16–1.40], 1.31 [1.18–1.45], 1.35 [1.22–1.49], respectively); leucine was not measured in EPIC-Potsdam. When combining isoleucine, tyrosine, and phenylalanine, the RR of T2D from lowest to highest quartile was 1.0, 1.13 (0.82–1.54), 1.45 (1.07–1.95), and 2.18 (1.62–2.95), respectively (P for trend < 0.0001).
Serum metabolites and risk of type 2 diabetes

We conducted several sensitivity analyses. In EPIC-Potsdam, a small proportion of the participants (14.3%) had fasted. The proportion of incident T2D cases was equally distributed among fasting and nonfasting participants (26.8% and 26.7%, respectively). Additional adjustment for fasting status did not change the results. Further, we did not observe an effect modification of fasting status on the association between metabolite factors 1 and 2 and T2D risk (P for interaction = 0.115 and 0.688, respectively). We observed no interaction with sex (P = 0.407 and 0.441, respectively), and in both men and women, the risk associations were similar. Hazard rates of T2D in different periods of follow-up were not different (factor 1 P = 0.126, factor 2 P = 0.794). Indicating that follow-up time did not affect the association between metabolite factors and risk of T2D. To ensure that the metabolite changes preceded the onset of T2D and were not attributed to prediabetic conditions, we repeated the analysis, excluding all cases of T2D that occurred shortly after the baseline examination.

TABLE 3
RR of T2D by quintiles of metabolite factors

| Factor 1† | EPIC-Potsdam | Replication in KORA |
|-----------|--------------|---------------------|
|           | Cases (N)    | RR (95% CI)*        | Cases (N)    | RR (95% CI)*        |
| Quartile 1| 296          | 1.00                | 37          | 1.00                |
| Quartile 2| 194          | 0.61 (0.46–0.80)    | 15          | 0.59 (0.31–1.33)    |
| Quartile 3| 148          | 0.50 (0.37–0.67)    | 20          | 0.60 (0.31–1.14)    |
| Quartile 4| 95           | 0.37 (0.27–0.51)    | 12          | 0.56 (0.27–1.21)    |
| Quartile 5| 67           | 0.31 (0.21–0.44)    | 7           | 0.49 (0.17–1.25)    |

P for trend

2.95 × 10⁻²⁰¹

1.00 × 10⁻¹

Data are partial Spearman correlation coefficients. Insulin sensitivity was adjusted for age and sex; insulin secretion was adjusted for age, sex, and insulin sensitivity. Insulin sensitivity and insulin secretion were estimated from OGTTs in the TuF study (n = 76). a, acyl-acyl; diacyl; di-acyl-alkyl; acyl-lysoPC; SM, sphingomyelin. †Relative risks were calculated with multivariate Cox regression (using the Prentice method to account for the case-cohort design in EPIC-Potsdam) across quintiles of metabolite factors after standardizing metabolite concentrations to a mean of 0 (SD 1). The model was adjusted for age, sex, alcohol intake from beverages (non-consumers; women >0–6, 6–12, and >12 g/day; men >0–12, 12–24, and >24 g/day), smoking (never, former, current <20 cigarettes/day, current ≥20 cigarettes/day), physical activity (cycling and sports in h/week), education (low, medium, high), coffee intake (cups/day), red meat intake (g/day), whole-grain bread intake (g/day), prevalent hypertension (yes/no), BMI (kg/m²), and waist circumference (cm). By conducting a PCA with orthogonal varimax rotation of all metabolites associated with T2D risk in EPIC-Potsdam, two metabolite factors could be identified that explained ≥50% of the variation of the metabolites. The corresponding linear factor equations (which equal the summed products of each metabolite’s standardized concentration and corresponding factor loading) were as follows: Factor 1 = (0.82 × PC ae C32:1) + (0.78 × PC ae C32:2) + (0.70 × PC ae C34:2) + (0.72 × PC ae C36:3) + (0.71 × PC ae C34:3) + (0.71 × PC ae C36:2) + (0.71 × PC ae C36:3) + (0.65 × PC ae C40:5) + (0.76 × PC ae C40:6) + (0.82 × PC ae C42:3) + (0.85 × PC ae C42:4) + (0.83 × PC ae C44:4) + (0.83 × PC ae C44:5) + (0.79 × PC ae C42:4) + (0.05 × SM C16:1) + (0.57 × SM OH C22:2) + (0.41 × lysoPC a C17:0) + (0.55 × propionyllcarnitine) + (0.66 × phenylalanine) + (0.61 × tryptophan) + (0.66 × tyrosine) + (0.68 × valine) + (0.66 × isoleucine) + (0.59 × PC aa C32:1) + (0.70 × PC aa C36:3) + (0.76 × PC aa C38:3) + (0.72 × PC aa C40:4) + (0.71 × PC aa C40:5) + (0.44 × hexose).

During the first 2 years of follow-up (n = 208). The risk associations were slightly lower, but not markedly different.

DISCUSSION
In this prospective investigation using a targeted metabolomic approach at population level, we found increased concentrations of hexose; phenylalanine; and diacyl-phosphatidylcholines C32:1, C36:1, C38:3, and C40:5 and reduced concentrations of glycine; sphingomyelin C16:1; acyl-alkyl-phosphatidylcholines C34:3, C40:6, C42:5, C44:4, and C44:5; and lysophosphatidylcholine C18:2 to be independently predictive of T2D in EPIC-Potsdam. The results agree with data from cross-sectional studies showing that patients with T2D had increased concentrations of sugar metabolites (13), acylcarnitines (14), and BCAA (13) and reduced concentrations of glycine; sphingomyelin C16:1; acyl-alkyl-phosphatidylcholines C34:3, C40:6, C42:5, C44:4, and C44:5; and lysophosphatidylcholine C18:2 to be independently predictive of T2D in EPIC-Potsdam. The results agree with data from cross-sectional studies showing that patients with T2D had increased concentrations of sugar metabolites (13), acylcarnitines (14), and BCAA (13) and reduced concentrations of glycine; sphingomyelin C16:1; acyl-alkyl-phosphatidylcholines C34:3, C40:6, C42:5, C44:4, and C44:5; and lysophosphatidylcholine C18:2 to be independently predictive of T2D in EPIC-Potsdam.
for the glucose-alanine cycle in skeletal muscle. Through alanine aminotransferase-catalyzed transamination reactions, this may result in increased substrate availability for hepatic gluconeogenesis, thereby increasing hepatic glucose production (39). Conversely, glycine is a gluconegenic amino acid; therefore, reduced serum glycine may also reflect increased gluconeogenesis. Alternative theories suggest that glycine depletion may reflect glutathione consumption driven by oxidative stress (40) or abundance of incompletely oxidized fuels that are excreted as urinary acylglycine conjugates (41–43). The frequently observed increase of BCAA in subjects with insulin resistance is also believed to be the result of reduced activities of key BCAA catabolic enzymes in liver and adipose tissue (44). Furthermore, amino acids may directly cause muscular insulin resistance by disrupting insulin signaling (45).

The positive association between hexose and T2D risk remained significant after adjustment for glucose. This observation could be an artifact from the different methods used to measure hexose and glucose. However, it has to be noted that hexose represented not only glucose but also the sum of all six-carbon monosaccharides. Previous studies have shown that in addition to glucose, fructose levels were elevated in individuals with T2D (12) and that intake of fructose was positively associated with risk of insulin resistance and T2D (46). In insulin-resistant conditions, the body aims to compensate for decreased glucose uptake of peripheral tissues through increased pancreatic insulin secretion (1). However, at the stage of overt insulin resistance, this system will eventually be exhausted, as caused by β-cell dysfunction and, subsequently, insulin secretion decreases (1). Phenylalanine, which was positively correlated to insulin secretion in the present study, may be involved in pathways to compensate early stages of insulin resistance through stimulation of insulin secretion. In contrast, increased hexose concentrations may indicate manifest insulin resistance and defect of β-cells. It is noteworthy that we observed significant associations between choline-containing phospholipids (i.e., diacyl-, acyl-alkyl-, and lysophosphatidylcholines and sphingomyelins) and T2D risk. Diacyl-phosphatidylcholines consist of glycerol linked to phosphocholine and two fatty acid residues, and removal of one fatty acid produces lysophosphatidylcholines. The corresponding acyl-alkyl-phosphatidylcholines comprise an ether linkage to one alkyl chain and one polyunsaturated fatty acid (47). Sphingomyelins are built of a ceramide core linked to one fatty acid and a phosphocholine or phosphoethanolamine (Fig. 3). Together, these phospholipids make up the main constituent of cellular membranes and may be involved in cellular signal transduction (48). In addition, they represent a major fraction of the human plasma lipidome because they are most abundant in all lipoproteins (49). Diacyl-phosphatidylcholines are particularly essential for hepatic secretion of triglyceride-rich VLDL particles and HDL (48), whereas acyl-alkyl-phosphatidylcholines may act as serum antioxidants to prevent lipoprotein oxidation (50). Their hepatic synthesis requires dietary choline (48). It was previously shown that choline-deficient mice on a high-fat diet showed reduced phosphatidylcholine biosynthesis and accumulated hepatic fat, but at the same time, they had reduced fasting insulin and improved glucose tolerance (51). In addition, impaired hepatic phosphatidylcholine biosynthesis led to reduced levels of plasma triglycerides and HDL cholesterol in vivo (52).
Accordingly, phosphatidylcholines and sphingomyelins were positively related to plasma HDL cholesterol in the present study. Furthermore, acyl-alkyl-phosphatidylcholines were inversely correlated to plasma triglycerides, opposite to diacyl-phosphatidylcholines, and higher levels of acyl-alkyl-phosphatidylcholines but not diacyl-phosphatidylcholines were linked to improved insulin sensitivity and reduced insulin secretion. Previous studies reported that acyl-alkyl-phosphatidylcholine levels were lower in obese subjects and subjects with insulin resistance (50,53). These mechanisms may contribute to the antithetical association between two phosphatidylcholine subclasses and T2D risk found in the present study and may indicate a key role of the type of linkage between phospholipid core and fatty acid residue. Furthermore, those phosphatidylcholines containing fatty acids with a lower number of carbons and double bonds were positively associated with T2D risk, contrary to those with a higher number of carbons and double bonds. Similar observations have recently been reported for fatty acid compositions of erythrocyte membrane phospholipids (54) and triglycerides (55), suggesting that lipids with a shorter chain length and saturated fatty acid residues may trigger development of T2D, whereas those containing longer chains and unsaturated fatty acids may offer protection.

In summary, the present data suggest that the identified metabolites could be part of different pathways involved in the early genesis of T2D. Therefore, these novel candidates could be useful in clinical practice to identify high-risk individuals earlier in order to delay or prevent disease onset. Furthermore, serum metabolites in this study predicted risk of T2D in a similar manner to a combination of classic risk factors; thus, their measurement may be a useful approach to predict T2D risk for individuals in the future. The metabolites could also serve as markers for specific metabolic pathways that are deranged and, thereby, allow the implementation of individualized preventive and therapeutic strategies. However, future investigations are warranted to calculate in detail the individual risks and to better understand the metabolic effects of these biomarkers and their biological mechanisms.

The primary strength of this study is that, to our knowledge, we were among the first to adopt a targeted metabolomic approach at population level and included a large sample from three independent, well-described study populations. Furthermore, our targeted metabolomic platform covered a wide variety of metabolites with known identity and quantitative measurements. Because we used a prospective design with consecutive follow-up,
we were able to investigate time-dependent exposure-disease associations.

The study, however, had several limitations. First, because we used independent study populations, the conditions of biosample collection, storage, and preparation were not necessarily the same, which may be a source of variation. In the KORA and TüF studies, fasting blood samples were collected from all participants, whereas in EPIC-Potsdam only a small proportion of participants provided fasting blood samples. Nevertheless, we did not observe an effect modification of fasting status on the association between metabolite factors and T2D risk, and we could reproduce very similar metabolite factors comparing fasting to nonfasting samples. Furthermore, the metabolomic analyses were based on serum samples in the EPIC-Potsdam and KORA studies but on plasma samples in the TüF study. As previously reported (56), the correlation between these serum and plasma metabolites was high; however, the absolute metabolite concentrations were higher in serum, which could lead to systematic changes. Second, the analytical method detected most of the metabolites with high specificity; however, it may not have detected all possible interferences among metabolites. Of the metabolites that we identified, diacylphosphatidylcholine C38:3 and sphingomyelin C16:1 may be interfering compounds for sphingomyelin C24:1 and diacylphosphatidylcholine C30:2, respectively. Third, we only had a limited number of incident T2D cases available from KORA. We may not have had sufficient statistical power, and the replication results have to be interpreted with caution. Fourth, there is a chance that reverse causation may explain the results, implying that overt diabetic conditions that were undiagnosed may have caused these metabolite changes. When we accounted for this issue, the results remained robust. Last, because this was an observational study, we cannot prove causality but only show associations. However, the identified metabolites were also correlated to established T2D biomarkers as well as to measures of insulin sensitivity and secretion in a different population, which underlines the biological plausibility of the results.

In conclusion, this prospective investigation using metabolomic data of independent study populations identified sugar metabolites, amino acids, and choline-containing phospholipids to be independently associated with risk of T2D. Beyond the classic pathways, these candidates point toward a novel role of phospholipid and lipoprotein metabolism in T2D pathophysiology. Future studies should further elucidate the biological mechanisms.

ACKNOWLEDGMENTS

The current study was supported by a grant from the Federal Ministry of Education and Research, Germany (Bundesministerium für Bildung und Forschung), to the German Center for Diabetes Research (Förderkennzeichen 01GI9222 to D.Z.D.). N.S. is supported by a Heisenberg professorship from the Deutsche Forschungsgemeinschaft.

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

A.Fl., H.B., and T.P. designed the study. A.Fl., N.S., Z.Y., and K.M. analyzed the data. A.Fl., N.S., Z.Y., K.M., D.D., H.-G.J., A.Fr., H.-U.H., M.H.A., A.P., M.R., C.P., R.W.-S., T.I., M.B.S., J.A., H.B., and T.P. provided their expertise and contributed to the writing of the manuscript. A.Fr., A.P., T.I., and H.B. collected the data. C.P. and J.A. conducted and supervised the metabolomic measurements. All authors take full responsibility for the contents of the manuscript. A.Fl. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

The authors thank all the study participants of EPIC-Potsdam, KORA, and TüF. Special thanks go to Sven Knüppel and Wolfgang Bernigau (Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke) for statistical advice and to Martin Floegel (Max Born Institute for Non-Linear Optics, Berlin) for his support in figure formatting.

REFERENCES

1. DeFronzo RA. Banting Lecture. From the triumvirate to the ominous octet: a new paradigm for the treatment of type 2 diabetes mellitus. Diabetes 2009;58:773–795
2. Peters AL, Davidson MB, Schriber DL, Hasselblad V. Meta-analysis Research Group on the Diagnosis of Diabetes Using Glycated Hemoglobin Levels. A clinical approach for the diagnosis of diabetes mellitus: an analysis using glycosylated hemoglobin levels. JAMA 1996;276:1246–1252
3. Herder C, Baumert J, Zierer A, et al. Immunological and cardiometabolic risk factors in the prediction of type 2 diabetes and coronary events: MONICA/KORA Augsburg case-cohort study. PLoS ONE 2011;6:e19852
4. Abu-Qamar M, Wilson A. Evidence-based decision-making: the case for diabetes care. Int J Evid Based Healthc 2007;5:254–260
5. Swedman M, Sayed Mahmoud M and Abdel-Fatah Ali A. Clinical implications of adiponectin and inflammatory biomarkers in type 2 diabetes mellitus. Dis Markers 2009;27:269–278
6. Li S, Shin HJ, Ding EL, van Dam RM. Adiponectin levels and risk of type 2 diabetes: a systematic review and meta-analysis. JAMA 2009;302:170–188
7. Ford ES, Schulze MB, Bergmann MM, Thamer C, Joost HG, Boeing H. Liver enzymes and incident diabetes: findings from the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study. Diabetes Care 2009;32:1139–1143
8. Stefan N, Fritsche A, Weikert C, et al. Plasma fetuin-A levels and the risk of type 2 diabetes. Diabetes 2008;57:2762–2767
9. Bain JR, Stevens RD, Wenner BR, Iiyama O, Muoio DM, Newgard CB. Metabolomics applied to diabetes research: moving from information to knowledge. Diabetes 2008;58:2429–2443
10. Newgard CB, An J, Bain JR, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. Cell Metab 2009;9:311–326
11. Wopereis S, Rubingh CM, van Erk MJ, et al. Metabolic profiling of the response to an oral glucose tolerance test detects subtle metabolic changes. PLoS ONE 2009;4:e5425
12. Fiehn O, Garvey WT, Newman JW, Lok KH, Hoppel CL, Adams SH. Plasma metabolomic profiles reflective of glucose homeostasis in non-diabetic and type 2 diabetic obese African-American women. PLoS ONE 2010;5:e15224
13. Suhrke K, Meisinger C, Döring A, et al. Immunological and cardiometabolic risk factors in the prediction of type 2 diabetes and coronary events: MONICA/KORA Augsburg case-cohort study. PLoS ONE 2011;6:e19852
14. Adams SH, Hoppel CL, Lok KH, et al. Plasma acylcarnitine profiles suggest incomplete long-chain fatty acid beta-oxidation and altered tricarboxylic acid cycle activity in type 2 diabetic African-American women. J Nutr 2009;139:1073–1081
15. Hu FB. Metabolic profiling of diabetes: from black-box epidemiology to systems epidemiology. Clin Chem 2011;57:1224–1226
16. Schulze MB, Hoffmann K, Boehling H, et al. An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. Diabetes Care 2007;30:510–515
17. Boeing H, Korfmann A, Bergmann MM. Recruitment procedures of EPIC-Germany. European Investigation into Cancer and Nutrition. Ann Nutr Metab 1999;43:205–215
18. Kroke A, Bergmann MM, Lotze G, Jeckel A, Klipstein-Grobusch K, Boeing H. Measures of quality control in the German component of the EPIC study. European Prospective Investigation into Cancer and Nutrition. Ann Nutr Metab 1999;43:216–224
SERUM METABOLITES AND RISK OF TYPE 2 DIABETES

19. Boeing H, Wahrenbrock J, Becker N. EPIC-Germany—a source for studies into diet and risk of chronic diseases. European Investigation into Cancer and Nutrition. Ann Nutr Metab 1999;43:195–204

20. Weikert C, Stefan N, Schulze MB, et al. Plasma fetuin-A levels and the risk of myocardial infarction and ischemic stroke. Circulation 2008;118:2555–2562

21. Montonen J, Drogan D, Joost HG, et al. Estimation of the contribution of biomarkers of different metabolic pathways to risk of type 2 diabetes. Eur J Epidemiol 2011;26:29–38

22. Bergmann MM, Bussas U, Boeing H. Follow-up procedures in EPIC-Germany—data quality aspects. European Prospective Investigation into Cancer and Nutrition. Ann Nutr Metab 1999;43:225–234

23. Schienkiewitz A, Schulze MB, Hoffmann K, Kroke A, Boeing H. Body mass index history and risk of type 2 diabetes: results from the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study. Am J Clin Nutr 2006;84:427–433

24. Rothman KJ, Greenland S, Lash TL. Case-control studies. In Modern epidemiology, 3rd ed. Rothman KJ, Greenland S, Eds. Philadelphia, Lippincott-Williams & Wilkins, 2008, p. 111–127

25. Meisinger C, Strassburger K, Heier M, et al. Prevalence of undiagnosed diabetes and impaired glucose regulation in 35-59-year-old individuals in Southern Germany: the KORA F4 Study. Diabet Med 2010;27:360–362

26. Rathmann W, Strassburger K, Heier M, et al. Incidence of type 2 diabetes in the elderly German population and the effect of clinical and lifestyle risk factors: KORA S4/F4 cohort study. Diabet Med 2009;26:1212–1219

27. Kowall B, Rathmann W, Strassburger K, Meisinger C, Holle R, Mielck A. Socioeconomic status is not associated with type 2 diabetes incidence in an elderly population in Germany: KORA S4/F4 cohort study. J Epidemiol Community Health 2011;65:606–612

28. Stefan N, Kantartzis K, Machann J, et al. Identification and characterization of metabolically benign obesity in humans. Arch Intern Med 2008;168:1065–1070

29. Matsuda M, DeFronzo RA. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. Diabetes Care 1999;22:1462–1470

30. Römisch-Margl W, Prehn C, Bogumil R, Röhring C, Suhre K, Adamski J. Ethanol metabolites and risk of type 2 diabetes: results from the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study. Am J Clin Nutr 2006;84:340–344

31. Prentice RL. Design issues in cohort studies. Stat Methods Med Res 1995;4:273–292

32. Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat 1979;6:65–70

33. Hardy RJ, Thompson SG. Detecting and describing heterogeneity in meta-analysis. Stat Med 1998;17:841–856

34. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557–560

35. Hsieh FY, Lavori PW. Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates. Control Clin Trials 2000;21:552–560

36. Hsieh FY, Lavori PW. Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates. Control Clin Trials 2000;21:552–560

37. Schienkiewitz A, Schulze MB, Hoffmann K, Kroke A, Boeing H. Body mass index history and risk of type 2 diabetes: results from the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study. Am J Clin Nutr 2006;84:427–433

38. Schmidt-Wilhelmsen KL. Serum metabolite profiling as a source for studies into diet and risk of chronic diseases. European Investigation into Cancer and Nutrition. Ann Nutr Metab 1999;43:225–234
**Supplementary Table 1.** Selection of metabolites according to their association with type 2 diabetes risk in EPIC-Potsdam.

| Rank* | Metabolite           | Mean (SD) µM† | α-Coefficient (SE)‡ | P-Value‡ | Corrected P-Value§ |
|-------|----------------------|---------------|---------------------|----------|--------------------|
| 1     | Hexose               | 4698 (984)    | 0.86 (0.07)         | 1.00 x10⁻³⁴ | 1.28 x10⁻³²*       |
| 2     | PC ae C34:3          | 8.51 (2.48)   | -0.45 (0.06)        | 1.81 x10⁻¹² | 2.28 x10⁻¹⁰*       |
| 3     | PC ae C44:6          | 1.18 (0.32)   | -0.40 (0.06)        | 5.70 x10⁻¹¹ | 7.13 x10⁻⁹*        |
| 4     | PC aa C42:0          | 0.60 (0.18)   | -0.36 (0.06)        | 1.01 x10⁻⁹  | 1.25 x10⁻⁷*        |
| 5     | PC aa C38:3          | 55.4 (14.8)   | 0.31 (0.05)         | 1.13 x10⁻⁹  | 1.39 x10⁻⁷*        |
| 6     | PC aa C42:1          | 0.30 (0.08)   | -0.34 (0.06)        | 2.69 x10⁻⁹  | 3.29 x10⁻⁷*        |
| 7     | PC ae C42:5          | 2.35 (0.54)   | -0.35 (0.06)        | 5.83 x10⁻⁹  | 7.06 x10⁻⁷*        |
| 8     | Phenylalanine        | 56.4 (12.0)   | 0.30 (0.05)         | 7.50 x10⁻⁹  | 9.01 x10⁻⁷*        |
| 9     | PC ae C44:5          | 1.77 (0.48)   | -0.35 (0.06)        | 1.82 x10⁻⁸  | 2.16 x10⁻⁶*        |
| 10    | PC ae C42:4          | 0.94 (0.24)   | -0.34 (0.06)        | 3.88 x10⁻⁸  | 4.58 x10⁻⁶*        |
| 11    | Isoleucine           | 211 (63.3)    | 0.26 (0.05)         | 2.60 x10⁻⁷  | 3.04 x10⁻⁵*        |
| 12    | Tyrosine             | 81.5 (22.7)   | 0.27 (0.05)         | 4.26 x10⁻⁷  | 4.94 x10⁻⁵*        |
| 13    | PC ae C40:5          | 3.99 (0.87)   | -0.28 (0.06)        | 4.45 x10⁻⁷  | 5.12 x10⁻⁵*        |
| 14    | Valine               | 294 (74.5)    | 0.24 (0.05)         | 5.17 x10⁻⁷  | 5.89 x10⁻⁵*        |
| 15    | Glycine              | 256 (77.2)    | -0.32 (0.06)        | 9.62 x10⁻⁷  | 1.09 x10⁻⁵*        |
| 16    | PC ae C32:2          | 0.72 (0.18)   | -0.28 (0.06)        | 9.62 x10⁻⁷  | 1.08 x10⁻⁵*        |
| 17    | lysoPC a C18:2       | 35.2 (13.7)   | -0.30 (0.06)        | 9.97 x10⁻⁷  | 1.11 x10⁻⁵*        |
| 18    | C3                   | 0.38 (0.14)   | 0.25 (0.05)         | 1.35 x10⁻⁶  | 1.49 x10⁻⁴*        |
| 19    | PC ae C44:4          | 0.38 (0.10)   | -0.28 (0.06)        | 2.70 x10⁻⁶  | 2.94 x10⁻⁴*        |
| 20    | PC ae C34:2          | 13.4 (3.68)   | -0.26 (0.06)        | 4.43 x10⁻⁶  | 4.79 x10⁻⁴*        |
| 21    | PC ae C40:6          | 5.52 (1.43)   | -0.26 (0.06)        | 5.45 x10⁻⁶  | 5.83 x10⁻⁴*        |
| 22    | SM C16:1             | 17.6 (3.86)   | -0.23 (0.05)        | 1.61 x10⁻⁵  | 1.71 x10⁻⁴*        |
| 23    | PC ae C32:1          | 2.85 (0.66)   | -0.23 (0.05)        | 2.28 x10⁻⁵  | 2.40 x10⁻⁴*        |
| 24    | PC ae C42:3          | 0.85 (0.21)   | -0.23 (0.06)        | 3.30 x10⁻⁵  | 3.44 x10⁻⁴*        |
| 25    | PC aa C40:4          | 4.00 (1.14)   | 0.19 (0.05)         | 6.05 x10⁻⁵  | 6.24 x10⁻⁴*        |
| 26    | SM (OH) C22:2        | 12.3 (3.43)   | -0.21 (0.05)        | 9.50 x10⁻⁵  | 9.69 x10⁻⁴*        |
| 27    | lysoPC a C17:0       | 1.98 (0.61)   | -0.22 (0.06)        | 1.73 x10⁻⁴  | 1.75 x10⁻²*        |
| 28    | Tryptophan           | 80.7 (12.6)   | 0.22 (0.06)         | 1.73 x10⁻⁴  | 1.73 x10⁻²*        |
| 29    | PC ae C36:3          | 9.57 (2.43)   | -0.21 (0.06)        | 2.01 x10⁻⁴  | 1.99 x10⁻²*        |
| 30    | PC aa C36:3          | 153 (36.6)    | 0.19 (0.05)         | 2.24 x10⁻⁴  | 2.20 x10⁻²*        |
| 31    | PC aa C32:1          | 16.9 (9.76)   | 0.17 (0.04)         | 2.56 x10⁻⁴  | 2.48 x10⁻²*        |
| 32    | PC ae C36:2          | 17.1 (4.52)   | -0.21 (0.06)        | 2.87 x10⁻⁴  | 2.76 x10⁻²*        |
| 33    | PC aa C40:5          | 11.4 (3.49)   | 0.16 (0.05)         | 3.60 x10⁻⁴  | 3.42 x10⁻²*        |
| 34    | PC aa C36:1          | 57.3 (15.6)   | 0.17 (0.05)         | 4.04 x10⁻⁴  | 3.79 x10⁻²*        |
| 35    | Glutamine            | 584 (94.5)    | -0.16 (0.05)        | 5.59 x10⁻⁴  | 5.20 x10⁻²*        |
| 36    | PC aa C42:2          | 0.21 (0.07)   | -0.19 (0.05)        | 6.32 x10⁻⁴  | 5.81 x10⁻²*        |
| 37    | C0                   | 36.6 (9.34)   | 0.17 (0.05)         | 1.01 x10⁻³  | 9.23 x10⁻²*        |
| 38    | PC ae C40:4          | 2.34 (0.54)   | -0.19 (0.06)        | 1.04 x10⁻³  | 9.40 x10⁻²*        |
### SUPPLEMENTARY DATA

|   |  |   |   |   |   |
|---|---|---|---|---|---|
|   |   |   |   |   |   |
| 39 | PC aa C38:4 | 116 (30.1) | 0.15 (0.05) | 1.50 x 10^{-3} | 1.33 x 10^{-4} |
| 40 | PC ae C38:6 | 8.41 (2.23) | -0.16 (0.05) | 1.50 x 10^{-3} | 1.32 x 10^{-4} |
| 41 | lysoPC a C18:1 | 19.5 (5.79) | -0.18 (0.06) | 1.60 x 10^{-3} | 1.39 x 10^{-4} |
| 42 | Arginine | 106 (23.0) | 0.17 (0.05) | 1.69 x 10^{-3} | 1.46 x 10^{-4} |
| 43 | PC aa C34:1 | 237 (63.1) | 0.16 (0.05) | 2.07 x 10^{-3} | 1.76 x 10^{-4} |
| 44 | PC ae C36:5 | 11.6 (3.23) | -0.16 (0.05) | 2.10 x 10^{-3} | 1.77 x 10^{-4} |
| 45 | SM C16:0 | 115 (23.8) | -0.15 (0.05) | 3.19 x 10^{-3} | 2.58 x 10^{-4} |
| 46 | Arginine | 106 (23.0) | 0.17 (0.05) | 1.69 x 10^{-3} | 1.46 x 10^{-4} |
| 47 | PC aa C34:1 | 237 (63.1) | 0.16 (0.05) | 2.07 x 10^{-3} | 1.76 x 10^{-4} |
| 48 | PC ae C40:3 | 1.16 (0.25) | -0.16 (0.06) | 3.48 x 10^{-3} | 2.78 x 10^{-4} |
| 49 | SM (OH) C14:1 | 7.45 (2.11) | -0.16 (0.06) | 3.79 x 10^{-3} | 3.00 x 10^{-4} |
| 50 | C5-DC / C6-OH | 0.03 (0.01) | 0.14 (0.05) | 4.58 x 10^{-3} | 3.57 x 10^{-4} |
| 51 | Methionine | 29.1 (7.90) | 0.14 (0.05) | 4.78 x 10^{-3} | 3.68 x 10^{-4} |
| 52 | PC aa C30:0 | 5.53 (1.77) | 0.15 (0.05) | 5.48 x 10^{-3} | 4.17 x 10^{-4} |
| 53 | C3-DC-M / C5-OH | 0.03 (0.01) | 0.13 (0.05) | 6.59 x 10^{-3} | 4.94 x 10^{-4} |
| 54 | SM C24:1 | 49.8 (14.1) | -0.14 (0.05) | 8.01 x 10^{-3} | 6.46 x 10^{-4} |
| 55 | PC ae C40:1 | 1.49 (0.39) | -0.14 (0.05) | 8.85 x 10^{-3} | 6.46 x 10^{-4} |
| 56 | PC aa C40:6 | 32.8 (10.9) | 0.12 (0.05) | 9.31 x 10^{-3} | 6.71 x 10^{-4} |
| 57 | PC aa C38:0 | 3.21 (0.91) | -0.14 (0.05) | 9.49 x 10^{-3} | 6.74 x 10^{-4} |
| 58 | lysoPC a C14:0 | 5.68 (0.97) | 0.13 (0.05) | 1.16 x 10^{-2} | 8.15 x 10^{-4} |
| 59 | PC aa C36:2 | 278 (58.7) | 0.12 (0.05) | 1.43 x 10^{-2} | 9.89 x 10^{-4} |
| 60 | C8:1 | 0.11 (0.05) | 0.13 (0.05) | 1.48 x 10^{-2} | 1 |
| 61 | PC ae C38:4 | 14.3 (3.24) | -0.13 (0.05) | 1.66 x 10^{-2} | 1 |
| 62 | PC aa C42:5 | 0.43 (0.14) | 0.12 (0.05) | 1.67 x 10^{-2} | 1 |
| 63 | SM C18:1 | 12.0 (3.19) | -0.13 (0.06) | 1.79 x 10^{-2} | 1 |
| 64 | PC aa C36:4 | 221 (56.7) | 0.13 (0.05) | 2.22 x 10^{-2} | 1 |
| 65 | Ornithine | 100 (27.2) | 0.12 (0.05) | 2.40 x 10^{-2} | 1 |
| 66 | PC aa C34:2 | 434 (93.4) | 0.12 (0.05) | 2.44 x 10^{-2} | 1 |
| 67 | SM (OH) C16:1 | 3.87 (1.12) | -0.12 (0.05) | 3.12 x 10^{-2} | 1 |
| 68 | PC ae C38:0 | 2.33 (0.75) | -0.11 (0.05) | 3.29 x 10^{-2} | 1 |
| 69 | C16 | 0.12 (0.03) | 0.11 (0.05) | 3.72 x 10^{-2} | 1 |
| 70 | PC aa C32:3 | 0.60 (0.16) | -0.11 (0.06) | 5.06 x 10^{-2} | 1 |
| 71 | PC ae C30:1 | 0.41 (0.20) | -0.12 (0.06) | 5.39 x 10^{-2} | 1 |
| 72 | C10 | 0.25 (0.13) | 0.07 (0.04) | 6.13 x 10^{-2} | 1 |
| 73 | SM C26:0 | 0.23 (0.08) | -0.09 (0.05) | 7.23 x 10^{-2} | 1 |
| 74 | PC aa C36:0 | 2.45 (0.82) | -0.09 (0.05) | 7.77 x 10^{-2} | 1 |
| 75 | SM C26:1 | 0.55 (0.18) | -0.09 (0.05) | 8.82 x 10^{-2} | 1 |
| 76 | PC aa C36:6 | 1.41 (0.55) | -0.08 (0.05) | 1.08 x 10^{-1} | 1 |
| 77 | C16:2 | 0.01 (0.006) | 0.08 (0.05) | 1.17 x 10^{-1} | 1 |
| 78 | lysoPC a C16:0 | 112 (25.5) | 0.08 (0.05) | 1.24 x 10^{-1} | 1 |
| 79 | C9 | 0.06 (0.02) | 0.07 (0.05) | 1.53 x 10^{-1} | 1 |
| 80 | SM C20:2 | 0.71 (0.41) | -0.08 (0.06) | 1.54 x 10^{-1} | 1 |
| 81 | PC ae C30:0 | 0.42 (0.13) | -0.08 (0.06) | 1.72 x 10^{-1} | 1 |

©2012 American Diabetes Association. Published online at http://diabetes.diabetesjournals.org/lookup/suppl/doi:10.2337/db12-0495/-/DC1
|    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|
| 82 | PC ae C38:2 | 2.19 (0.55) | -0.07 (0.05) | 1.79 x 10^{-01} | 1 |
| 83 | lysoPC a C16:1 | 3.41 (1.23) | 0.06 (0.05) | 1.80 x 10^{-01} | 1 |
| 84 | lysoPC a C20:4 | 6.37 (1.98) | -0.07 (0.06) | 1.87 x 10^{-01} | 1 |
| 85 | lysoPC a C20:3 | 2.48 (0.78) | 0.07 (0.05) | 1.94 x 10^{-01} | 1 |
| 86 | PC ae C40:2 | 2.25 (0.57) | -0.06 (0.05) | 2.31 x 10^{-01} | 1 |
| 87 | PC aa C40:3 | 0.59 (0.21) | 0.06 (0.05) | 2.55 x 10^{-01} | 1 |
| 88 | lysoPC a C20:4 | 6.37 (1.98) | -0.07 (0.06) | 1.87 x 10^{-01} | 1 |
| 89 | C10:2 | 0.05 (0.02) | 0.06 (0.05) | 2.61 x 10^{-01} | 1 |
| 90 | PC ae C30:2 | 0.14 (0.05) | -0.07 (0.06) | 2.77 x 10^{-01} | 1 |
| 91 | PC aa C28:1 | 3.64 (0.90) | -0.06 (0.05) | 2.83 x 10^{-01} | 1 |
| 92 | PC aa C42:6 | 0.72 (0.17) | 0.06 (0.05) | 2.92 x 10^{-01} | 1 |
| 93 | PC ae C34:3 | 18.5 (5.74) | 0.06 (0.05) | 3.10 x 10^{-01} | 1 |
| 94 | PC ae C34:1 | 10.2 (2.46) | -0.06 (0.06) | 3.13 x 10^{-01} | 1 |
| 95 | PC ae C42:2 | 0.70 (0.17) | -0.05 (0.05) | 3.22 x 10^{-01} | 1 |
| 96 | Threonine | 97.1 (29.2) | -0.05 (0.06) | 3.26 x 10^{-01} | 1 |
| 97 | PC aa C38:1 | 0.68 (0.43) | -0.05 (0.06) | 3.48 x 10^{-01} | 1 |
| 98 | PC ae C36:4 | 19.1 (5.10) | -0.05 (0.05) | 3.50 x 10^{-01} | 1 |
| 99 | PC ae C44:3 | 0.11 (0.03) | -0.05 (0.05) | 3.62 x 10^{-01} | 1 |
| 100 | Proline | 215 (71.0) | 0.04 (0.05) | 3.88 x 10^{-01} | 1 |
| 101 | SM C24:0 | 25.9 (6.71) | 0.04 (0.05) | 4.03 x 10^{-01} | 1 |
| 102 | C18:2 | 0.07 (0.02) | -0.04 (0.05) | 4.17 x 10^{-01} | 1 |
| 103 | PC aa C34:4 | 2.36 (0.87) | 0.05 (0.06) | 4.20 x 10^{-01} | 1 |
| 104 | PC ae C36:0 | 0.80 (0.27) | 0.04 (0.05) | 4.41 x 10^{-01} | 1 |
| 105 | PC aa C42:4 | 0.20 (0.05) | 0.04 (0.05) | 4.43 x 10^{-01} | 1 |
| 106 | SM (OH) C24:1 | 1.59 (0.48) | -0.03 (0.05) | 4.78 x 10^{-01} | 1 |
| 107 | PC ae C36:1 | 9.46 (2.32) | -0.04 (0.06) | 4.93 x 10^{-01} | 1 |
| 108 | C14:1 | 0.20 (0.05) | -0.03 (0.05) | 5.00 x 10^{-01} | 1 |
| 109 | C18:1 | 0.15 (0.05) | 0.03 (0.05) | 5.07 x 10^{-01} | 1 |
| 110 | PC ae C42:1 | 0.35 (0.09) | -0.03 (0.05) | 5.80 x 10^{-01} | 1 |
| 111 | PC aa C36:5 | 33.9 (17.2) | -0.03 (0.05) | 6.08 x 10^{-01} | 1 |
| 112 | PC ae C38:1 | 4.77 (1.14) | -0.03 (0.05) | 6.08 x 10^{-01} | 1 |
| 113 | PC aa C38:5 | 59.3 (16.6) | 0.02 (0.05) | 6.20 x 10^{-01} | 1 |
| 114 | C14:2 | 0.03 (0.01) | 0.02 (0.05) | 6.64 x 10^{-01} | 1 |
| 115 | SM (OH) C22:1 | 14.7 (3.96) | 0.02 (0.05) | 6.90 x 10^{-01} | 1 |
| 116 | PC ae C38:3 | 4.8 (1.14) | -0.02 (0.06) | 6.91 x 10^{-01} | 1 |
| 117 | Serine | 115 (28.4) | -0.02 (0.06) | 6.94 x 10^{-01} | 1 |
| 118 | SM C18:0 | 25.1 (6.21) | 0.02 (0.05) | 7.13 x 10^{-01} | 1 |
| 119 | lysoPC a C18:0 | 30.3 (8.12) | 0.02 (0.05) | 7.37 x 10^{-01} | 1 |
| 120 | Histidine | 93.4 (16.6) | -0.02 (0.06) | 7.80 x 10^{-01} | 1 |
| 121 | PC aa C40:2 | 0.38 (0.14) | -0.02 (0.06) | 7.85 x 10^{-01} | 1 |
| 122 | C7-DC | 0.04 (0.02) | 0.01 (0.05) | 8.01 x 10^{-01} | 1 |
| 123 | C2 | 7.03 (2.92) | 0.01 (0.05) | 8.32 x 10^{-01} | 1 |
| 124 | PC ae C34:0 | 1.85 (0.53) | 0.01 (0.06) | 8.32 x 10^{-01} | 1 |
**Supplementary Table 2.** Metabolites and their corresponding factor loadings as obtained from principal component analysis (PCA) comparing random and fasting blood sampling in EPIC-Potsdam.

| 125 | C18 | 0.05 (0.02) | 0.01 (0.05) | 8.67 x 10^-01 | 1 |
|-----|-----|-------------|-------------|---------------|---|
| 126 | PC aa C38:6 | 104 (31.0) | 0.00 (0.05) | 9.58 x 10^-01 | 1 |
| 127 | lysoPC a C28:1 | 0.62 (0.24) | 0.00 (0.06) | 9.67 x 10^-01 | 1 |

*Ranking from lowest to highest uncorrected p-value.
†Mean serum concentration in the EPIC-Potsdam subcohort n=2282.
‡β-coefficients and P-values were derived from multivariate Cox regression analysis modified using the prentice method to account for the case-cohort design in a continuous model after standardizing metabolite concentrations to mean of zero and standard deviation of one. The model was adjusted for age, sex, alcohol intake from beverages (non-consumers; women: >0-6 g/d, 6-12 g/d, >12 g/d; men: >0-12 g/d, 12-24 g/d, >24 g/d), smoking (never, former, current ≤20 cigarettes/d, current >20 cigarettes/d), physical activity (cycling and sports in h/week), education (no degree/vocational training; trade/technical school; university degree), coffee intake (cups/d), red meat intake (g/d), whole grain bread intake (g/d), prevalent hypertension (yes/no), BMI (kg/m²), and waist circumference (cm).
§P-values were corrected to account for multiple testing (n=127) using the Bonferroni-Holm procedure [1]. *Corrected p-value <0.05.

Abbreviations: a, acyl; aa, diacyl; ae, acyl-alkyl; DC, decarboxyl; EPIC, European Prospective Investigation into Cancer and Nutrition; PC, phosphatidylcholine; SE, standard error; SM, sphingomyelin.
| Metabolite          | 72* | 76* | 1  | 3  |
|--------------------|-----|-----|----|----|
| PC ae C34:3        | 71* | 78* | 15 | 8  |
| PC ae C36:2        | 71* | 76* | 20 | 17 |
| PC ae C36:3        | 85* | 87* | 15 | 18 |
| PC ae C40:5        | 76* | 76* | 11 | 12 |
| PC ae C40:6        | 82* | 82* | 14 | 14 |
| PC ae C42:3        | 85* | 85* | -7 | -4 |
| PC ae C42:4        | 87* | 88* | -7 | -5 |
| PC ae C42:5        | 76* | 78* | -1 | -1 |
| PC ae C44:4        | 78* | 77* | -7 | -4 |
| PC ae C44:5        | 83* | 82* | -14| -12|
| LysoPC ae C17:0    | 41* | 48* | 9  | 5  |
| LysoPC ae C18:2    | 29  | 32  | 6  | 7  |
| SM C16:1           | 54* | 57* | 26 | 32 |
| SM OH C22:2        | 57* | 61* | 10 | 13 |
| Hexose             | -5  | 2   | 44*| 62*|

*factor loading > 0.35.
†Metabolite factors were identified by PCA, based on the correlation matrix, of all metabolites significantly associated with risk of type 2 diabetes in the European Prospective Investigation into Cancer and Nutrition Potsdam study. An orthogonal varimax rotation was used, and two factors were retained as they accounted for more than 50% of the observed variance (see also Figure 1 in the manuscript).
‡Random blood: n= 2997, of them fasting blood: n=429.
**Supplementary Table 3.** Correlation* between metabolites associated with type 2 diabetes risk and established biomarkers in the EPIC-Potsdam subcohort

| Glucose† | ALT | GGT | HDL | Triglycerides† | Adiponectin | Hs-CRP | HbA1c | Fetuin A |
|---------|-----|-----|-----|---------------|-------------|--------|-------|---------|
| Hexose  | 0.56| 0.04| 0.05| -0.03         | 0.05        | -0.04  | 0.03  | 0.09    | 0.01    |
| Phenylalanine | 0.13| 0.08| 0.13| -0.02         | 0.05        | -0.04  | 0.05  | 0.04    | 0.02    |
| Glycine | -0.04| 0.00| -0.05| -0.02        | -0.06       | 0.06   | -0.20| -0.05   | -0.12   |
| SM C16:1 | 0.05| 0.01| 0.02| 0.27         | -0.03       | 0.16   | 0.01  | 0.06    | 0.05    |
| PC ae C34:3 | -0.14| -0.06| -0.08| 0.41        | -0.19       | 0.24   | -0.06| -0.04   | 0.00    |
| PC ae C40:6 | -0.13| 0.01| -0.03| 0.20         | -0.10       | 0.09   | -0.07| -0.09   | 0.04    |
| PC ae C42:5 | -0.15| -0.06| -0.06| 0.23        | -0.19       | 0.15   | -0.03| -0.12   | 0.00    |
| PC ae C44:4 | -0.15| -0.08| -0.06| 0.23        | -0.16       | 0.14   | -0.02| -0.09   | 0.02    |
| PC ae C44:5 | -0.10| -0.06| -0.03| 0.25         | -0.18       | 0.14   | 0.00  | -0.09   | -0.01   |
| PC aa C32:1 | 0.04| 0.10| 0.12| 0.21         | 0.35        | 0.08   | 0.12  | 0.03    | 0.08    |
| PC aa C36:1 | 0.03| 0.12| 0.11| 0.26         | 0.30        | 0.07   | -0.02| 0.05    | 0.01    |
| PC aa C38:3 | 0.00| 0.16| 0.18| 0.18         | 0.39        | -0.01  | 0.09  | 0.08    | 0.13    |
| PC aa C40:5 | 0.02| 0.15| 0.16| 0.16         | 0.37        | 0.00   | -0.06| 0.04    | 0.06    |
| lysoPC a C18:2 | 0.00| -0.06| -0.07| 0.09        | -0.07       | 0.04   | -0.21| -0.05   | -0.10   |
| Factor 1‡ | -0.20| -0.06| -0.07| 0.36        | -0.20       | 0.22   | -0.06| -0.12   | 0.02    |
| Factor 2§ | 0.14| 0.16| 0.17| 0.08         | 0.37        | -0.06  | 0.05  | 0.07    | 0.07    |

*Presented are partial Spearman correlation coefficients adjusted for age, sex, alcohol intake from beverages (non-consumers; women: >0-6 g/d, 6-12 g/d, >12 g/d; men: >0-12 g/d, 12-24 g/d, >24 g/d), smoking (never, former, current ≤20 cigarettes/d, current >20 cigarettes/d), physical activity (cycling and sports in h/week), education (low, medium, high), coffee intake (cups/d), red meat intake (g/d) and whole grain bread intake (g/d), prevalent hypertension (yes/no), BMI (kg/m²), and waist circumference (cm). n=2138.

†Based on fasting blood samples only: n=303.

‡Factor 1= 0.82 x PC aa C42:0 + 0.79 x PC aa C42:1 + 0.80 x PC ae C32:1 + 0.78 x PC ae C32:2 + 0.70 x PC ae C34:2 + 0.72 x PC ae C34:3 + 0.71 x PC ae C36:2 + 0.71 x PC ae C36:3 + 0.85 x PC ae C40:5 + 0.76 x PC ae C40:6 + 0.82 x PC ae C42:3 + 0.85 x PC ae C42:4 + 0.87 x PC ae C42:5 + 0.76 x PC ae C44:4 + 0.78 x PC ae C44:5 + 0.83 x PC ae C44:6 + 0.54 x SM C16:1 + 0.57 x SM OH C22:2 + 0.41 x lysoPC ae C18:2.

§Factor 2= 0.55 x propionyl-carnitine + 0.66 x phenylalanine + 0.61 x tryptophan + 0.66 x tyrosine + 0.68 x valine + 0.66 x isoleucine + 0.59 x PC aa C32:1 + 0.70 x PC aa C36:1 + 0.65 x PC aa C36:3 + 0.76 x PC aa C38:3 + 0.72 x PC aa C40:4 + 0.71 x PC aa C40:5 + 0.44 x hexose.
SUPPLEMENTARY DATA

Abbreviations: a, acyl; aa, diacyl; ae, acyl-alkyl; ALT, alanine transaminase; EPIC, European Prospective Investigation into Cancer and Nutrition; GGT, $\gamma$-glutamyltransferase; HbA$_{1c}$, glycated hemoglobin A1c; HDL, high density lipoprotein cholesterol; hs-CRP, high sensitivity C-reactive protein; PC, phosphatidylcholine; SM, sphingomyelin.

**Supplementary Table 4.** Measures* of discrimination and calibration comparing different multivariable adjusted models to predict type 2 diabetes in EPIC-Potsdam

| Model                              | -2LOG-Likelihood Ratio | Nagelkerke R$^2$† | Hosmer-Lemeshow‡ | Receiver Operating Characteristic (ROC) § |
|------------------------------------|------------------------|-------------------|------------------|------------------------------------------|
|                                    | 2413                   | 0.40              | 15.89            | 0.044                                    | 0.847 (0.832-0.862) | Ref. |
| Diabetes Risk Score (DRS)¶         | 2387                   | 0.41              | 13.53            | 0.095                                    | 0.849 (0.833-0.865) | 0.838 |
| Metabolites                        | 2210                   | 0.47              | 18.41            | 0.018                                    | 0.873 (0.859-0.887) | <0.0001 Ref. |
| DRS + Glucose                      | 2083                   | 0.52              | 9.39             | 0.310                                    | 0.890 (0.877-0.903) | <0.0001 <0.0001 |
| DRS + Metabolites                  | 1978                   | 0.55              | 15.25            | 0.054                                    | 0.901 (0.888-0.913) | <0.0001 Ref. |
| DRS + Glucose + HbA$_{1c}$         | 1864                   | 0.58              | 15.04            | 0.058                                    | 0.912 (0.901-0.924) | <0.0001 <0.0001 |

*Better model performance is indicated by smaller -2LOG-Likelihood Ratio, larger Nagelkerke’s R$^2$, good calibration with a Hosmer-Lemeshow P-Value ≥0.05, and better discrimination mirrored by larger C-statistics with significant improvement of model performance, P for difference <0.05.

†Specifically, Nagelkerke’s R$^2$ defines the fraction of -2 log likelihood explained by the predictors ranging from 0 to 1. Larger values indicate better model fit [2].

‡As a measure of model calibration, the Hosmer-Lemeshow statistic compares predicted and observed probabilities of type 2 diabetes derived from deciles of predicted risk. Smaller χ2 values and larger p-values specify better model fit. P-values <0.05 indicate significant difference between expected and observed probabilities [3].

§The ROC curve is a plot of sensitivity versus false-positive rate of a continuous prediction model on a binary outcome. The c-statistic equals the area under the ROC curve, a measure of discrimination that mirrors the probability the model assigns a higher risk to future type 2 diabetes cases compared to controls. It may range from 0.5 (no discrimination) to 1.0 (perfect discrimination).
The method of DeLong [5] was used to test for difference between c-statistic, \( p < 0.05 \) indicate significant model improvement.

The German Diabetes Risk Score [6] combines information on several diabetes risk factors, such as diet, lifestyle and anthropometry, to estimate risk of developing type 2 diabetes. The German Diabetes Risk Score is computed according to the following formula: 

\[
\text{German Diabetes Risk Score} = 7.4 \times \text{waist (cm)} - 2.4 \times \text{height (cm)} + 4.3 \times \text{age (years)} + 46 \times \text{hypertension (self-report)} + 49 \times \text{red meat (each 150 g/day)} - 9 \times \text{whole-grain bread (each 50 g/day)} - 4 \times \text{coffee (each 150 g/day)} - 20 \times \text{moderate alcohol (between 10 and 40 g/day)} - 2 \times \text{physical activity (h/week)} + 24 \times \text{former smoker} + 64 \times \text{current heavy smoker (≥20 cigarettes/day)}.
\]

Metabolites: hexose, phenylalanine, glycine, SM C16:1, PC aa C32:1, PC aa C36:1, PC aa C38:3, PC aa C40:5, PC ae C34:3, PC ae C40:6, PC ae C42:5, PC ae C44:4, PC ae C44:5, lysoPC a C18:2.

Abbreviations: a, acyl; aa, diacyl; ae, acyl-alkyl; CI, confidence interval; DRS, Diabetes Risk Score; EPIC, European Prospective Investigation into Cancer and Nutrition; PC, phosphatidylcholine; ROC, receiver operating characteristic; SM, sphingomyelin.