Non-targeted metabolomic biomarkers and metabotypes of type 2 diabetes: A cross-sectional study of PREDIMED trial participants

M. Urpi-Sarda [a,b,1,*], E. Almanza-Aguilera [a,b,1], R. Llorach a,b, R. Vazquez-Fresno [a,c], R. Estruch [d,e], D. Corella [e,f], J.V. Sorli [e,f], F. Carmona [g], A. Sanchez-Pla [g], J. Salas-Salvado´ [e,h], C. Andres-Lacueva [a,b,*]

[a] *Biomarkers and Nutrimetabolomic Laboratory, Department of Nutrition, Food Sciences and Gastronomy, XaRTA, INSA-UB, Faculty of Pharmacy and Food Sciences, University of Barcelona, Barcelona, Spain*

[b] *CIBER de Fragilidad y Envejecimiento Saludable (CIBERFES), Instituto de Salud Carlos III, Barcelona, Spain*

[c] *Department of Computing Science and Biological Sciences, University of Alberta, Edmonton, Canada*

[d] *Department of Internal Medicine, Hospital Clinic, Institut d'Investigacions Biome`diques August Pi Sunyer (IDIBAPS), Barcelona, Spain*

[e] *CIBER Pathophysiology of Obesity and Nutrition (CIBERobn), Instituto de Salud Carlos III, Madrid, Spain*

[f] *Department of Preventive Medicine and Public Health, University of Valencia, Valencia, Spain*

[g] *Statistics Department, Biology Faculty, University of Barcelona, Barcelona, Spain*

[h] *Human Nutrition Unit, Biochemistry and Biotechnology Department. Hospital Universitari de Sant Joan de Reus, Institut d'Investigacio Sanitaria Pere Virgili, Universitat Rovira i Virgili, Reus, Spain*

Keywords Metabolomics NMR Metabotypes Multi-metabolite signature PREDIMED Type 2 diabetes

Abbreviations: T2D, type 2 diabetes; CVD, cardiovascular disease; PREDIMED, Prevention with Mediterranean Diet; FID, free induction decay; OSC, orthogonal signal correction; PLS-DA, partial least squares discriminant analysis; VIP, variable importance in projection; AA, acetoacetate; ROC, receiver operating characteristic; AUROC, area under the ROC curve; PAG, phenylacetylglutamine.

* Corresponding authors at: Av. Joan XXIII 27-31, 08028 Barcelona, Spain. E-mail addresses: murpi@ub.edu (M. Urpi-Sarda), candres@ub.edu (C. Andres-Lacueva).

1 These authors contributed equally to this work.

ABSTRACT

Aim. – To characterize the urinary metabolomic fingerprint and multi-metabolite signature associated with type 2 diabetes (T2D), and to classify the population into metabotypes related to T2D. Methods. – A metabolomics analysis using the 1H-NMR-based, non-targeted metabolomic approach was conducted to determine the urinary metabolomic fingerprint of T2D compared with non-T2D participants in the PREDIMED trial. The discriminant metabolite fingerprint was subjected to logistic regression analysis and ROC analyses to establish and to assess the multi-metabolite signature of T2D prevalence, respectively. Metabotypes associated with T2D were identified using the k-means algorithm.

Results. – A total of 33 metabolites were significantly different ($P < 0.05$) between T2D and non-T2D participants. The multi-metabolite signature of T2D comprised high levels of methylsuccinate, alanine, dimethylglycine and guanidoacetate, and reduced levels of glutamine, methylguanidine, 3-hydro- xymandelate and hippurate, and had a 96.4% AUC, which was higher than the metabolites on their own and glucose. Amino-acid and carbohydrate metabolism were the main metabolic alterations in T2D, and various metabotypes were identified in the studied population. Among T2D participants, those with a metabotype of higher levels of phenylalanine, phenylacetylglutamine, p-cresol and acetoacetate had significantly higher levels of plasma glucose.

Conclusion. – The multi-metabolite signature of T2D highlights the altered metabolic fingerprint associated mainly with amino-acid, carbohydrate and microbiota metabolism. Metabotypes identified in this patient population could be related to higher risk of long-term cardiovascular events and therefore require further studies. Metabolomics is a useful tool for elucidating the metabolic complexity and interindividual variation in T2D towards the development of stratified precision nutrition and medicine. Trial registration at www.controlled-trials.com: ISRCTN35739639.

Introduction

Type 2 diabetes (T2D) encompasses individuals who have hyperglycaemia resulting from defects in insulin secretion, insulin action or both [1]. Moreover, hyperglycaemia and insulin resistance are risk factors for cardiovascular disease (CVD) [2]. Besides an understanding of the pathophysiology of T2D, the identification of individuals at high risk, as well as knowledge of the metabolic alterations produced in patients with T2D, are crucial for preventative and disease management strategies. In recent years, progress in the development of biomarkers for T2D has been achieved due to advances in the emerging '-omics' technologies, including metabolomics [3]. Successful applications of metabolomics in T2D research include the discovery of biomarkers for diagnoses and prognoses, altered metabolic pathways and drug mechanisms of action [4]. Currently, of the high-throughput analytical techniques, high-performance liquid chromatography–mass spectrometry (HPLC–MS) and proton nuclear magnetic resonance (1H-NMR) spectroscopy are those most widely employed in metabolomics for the study of diabetes, mostly due to their advantages in the analysis and identification of a broad range of metabolites in biofluids [3,4]. Furthermore, 1H-NMR is frequently used in non-targeted metabolomic approaches to profile metabolites in studies comparing T2D and non-T2D populations, as well as for elucidation and confirmation of the metabolic pathways altered as a consequence of T2D [3,5,6]. In recent years, several accurate prediction models have been constructed including variables such as age, gender and lifestyle factors [7], and some recent-omics technologies have the potential to serve as accurate analytical techniques for discovering novel biomarkers that could be involved in predictive models of T2D [4]. However, it should be pointed out that these models have mainly been tested with plasma samples, although a few have used urine samples and have also compared healthy vs T2D subjects. Thus, the use of these models in metabolomics is essential for identifying molecular signatures and phenotypic variations to improve prediction of disease risk and to better manage patients' care and outcomes [8]. Recently, the term 'stratified medicine' has emerged, based on the concept that some groups of individuals should be treated differently from others due to intervariability [8]. This inter-variability can be characterized by metabolomics through the study of metabolic phenotypes, or metabotypes, as the starting point for future stratified medicine programmes and lifestyle interventions [9]. In fact, patterns of variation or metabotypes have already been previously used in diabetes datasets to separate controls from patients

[10] or healthy from diseased groups [11], as well as for studying metabolomic differences among clinical phenotypes [12] and developing strategies for delivery of dietary advice [11,13]. In the present study, the aim was to characterize the urinary metabolomic fingerprint and multi-metabolite signature associat-ed with T2D prevalence in the Prevention with Mediterranean Diet (PREDIMED) study population, using a 1H-NMR-based, non-targeted metabolomic approach and classifying the population into metabolic phenotypes (metabotypes) in relation to T2D.

Methods

Study population

The PREDIMED was a parallel-group, single-blind, multicentre, randomized controlled 5-year clinical trial aimed at assessing the effects of the Mediterranean diet on CVD primary prevention [14]. Full details of the study design and protocol have been published elsewhere (www.predimed.es) [15]. Briefly, the parti-cipants were men (55–80 years of age) and women (60–80 years of age) with T2D and/or at least three of the following cardiovascular risk factors: hypertension; overweight [body mass index (BMI) > 25 kg/m2]; current smoker; low-density lipoprotein (LDL) cholesterol >4.14 mmol/L; high-density lipoprotein (HDL) cholesterol <1.03 mmol/L; and a family history of premature CVD. The trial was registered at www.controlled-trials.com as ISRCTN35739639. The present PREDIMED substudy used data collected from 154 consecutive participants from two centres (Hospital Clinic of Barcelona and University of Valencia) where urinary metabolome was determined using the 1H-NMR approach at baseline. Of these 154 participants, 85 were T2D patients and 69 were non-T2D subjects. The former were diagnosed as previously reported [14,16], and all participants were free of diabetic nephropathy. The institutional review boards of the two centres approved the study protocols, and written informed consent was given by all participants.

Urine collection and measurements

Spot urine samples were collected at baseline and immediately stored at -80ºC until analysis. Trained personnel performed the anthropometric and blood-pressure measurements. Validated questionnaires were employed to record physical activity, lifestyle, disease history and medication use [15].

Metabolomic analysis: 1H-NMR sample preparation, data acquisition and processing

Urine samples were thawed at 4º C and gently vortexed before metabolomic analysis, using a procedure based on a previously published methodology [17]. Briefly, 300 mL of urine were diluted in 200 mL of H2O/D2O (8:2 ratio) and mixed with an internal standard solution [0.1% chemical-shift reference 3-(trimethylsi-lyl)propionic-2,2,3,3-d4 acid sodium salt (TSP), 2 mM of sodium azide (NaN3) and 1.5 M of KH2PO4 in 99% deuterated water (D2O)]; the pH was set at 7.0 with a KOD solution. The 1H-NMR experiments were conducted using a 500-MHz spectrometer (Varian INOVA; Varian Medical Systems, Palo Alto, CA, USA), with presaturation of water resonance using a nuclear Overhauser enhancement (NOESY)-presat pulse sequence. Internal tempera-ture was kept constant at 298 K during acquisition. Spectra were acquired by collecting 128 scans at 32-K datapoints with a spectral width of 14 ppm, acquisition time of 2 s, relaxation delay of 5 s and mixing time of 100 ms. For spectral processing, the free induction decay (FID) was multiplied by an exponential function corres-ponding to a 0.3-Hz line broadening before Fourier transformation. All spectra were phased, baseline-corrected and referenced to TSP (d 0.0) using TopSpin version 3.2 software (Bruker BioSpin GmbH, Rheinstetten, Germany). The spectral data were processed, using an intelligent bucketing algorithm, in domains of 0.005 ppm [17] and integrated using ACD/NMR Processor 12.0 software (Advanced Chemistry Development, Inc., Toronto, ON, Canada). The spectral region 4.75–5.00 ppm was excluded from the dataset to avoid spectral interference from residual water.

Statistical analysis

A dataset containing integrals of NMR spectra was imported into MetaboAnalyst 3.0, a web-based platform for extensive analysis of metabolomic data [18], filtered using interquantile range (IQR) and row-wise-normalized by the sum of the spectral intensities. The normalized dataset was then imported into SIMCA-P+ 13.0 (Umetrics, Umea , Sweden) before being log-transformed and range-scaled prior to performing a principal component analysis to explore data distribution [17]. To reduce variability not associated with T2D classification, orthogonal signal correction (OSC) was applied to the dataset followed by partial least squares discriminant analysis (PLS-DA) to determine differences in metabolite profiling between the T2D and non-T2D groups. The predictive ability of the OSC–

PLS-DA models was then evaluated: one-third of the samples (validation set) were randomly removed from the whole dataset (training set), and the OSC–PLS-DA models calculated. This procedure was repeated five times, and was used to evaluate the ability of the models to classify prediction sets, and to calculate quality parameters of the method and the misclassification table. Quality and validation of the resultant OSC–PLS-DA models were assessed through R2Y(cum) and Q2(cum) parameters (calculated by seven-round internal data cross-validation using the default algorithm provided by the SIMCA-P+ 13.0 software), as well as by a permutation test (n = 200). Discriminant features between T2D patients and non- T2D subjects were identified from their variable importance in projection (VIP) values >1.0, a generally accepted threshold in metabolomic studies [17], with the VIP-sd(VIP) parameter also included as an additional quality parameter of the method. To eliminate the confusing effect of waist circumference, a general-ized linear model was applied, using the metabolic signature as an independent variable and T2D as a dependent variable, and adjusted by waist cirumference. Characteristics between participants were compared by apply-ing Student's t test and the chi-squared test for continuous and qualitative variables, respectively. Differences in metabolites between the T2D and non-T2D groups were tested using Student's t test with a Benjamini–Hochberg procedure for adjusting P values. The significance level was set at $P < 0.05$. Univariate analyses and generalized linear models were performed using IBM SPSS version 21 software (IBM Corp., Armonk, NY, USA).

Metabolite identification

Identification of metabolites was achieved using the Chenomx NMR Suite 7.6 Profiler (Chenomx Inc, Edmonton, AB, Canada). In addition, NMR spectral libraries were consulted in databases such as the Human Metabolome Database [19] and Biological Magnetic Resonance Data Bank [20], together with the currently available literature on NMR-based metabolomics [4,17].

Metabolic pathway analysis

Identified metabolites were submitted to the Pathway Analysis and Network Analysis modules in MetaboAnalyst 3.0 [18] and MetaCoreTM (GeneGo, Inc., St. Joseph, MI, USA), respectively, to undergo analyses of metabolic pathways and biological inter-pretations of metabolites related to T2D.

Multi-metabolite signature model for T2D prevalence

The results obtained by OSC–PLS-DA analysis were subjected to forward conditional stepwise logistic regression analysis to design a multi-metabolite signature model of T2D prevalence. The prediction model was applied to a training set (two-thirds of participants) and subsequently validated against a validation set (one-third of participants). Quality of the models was evaluated by calculating the sensitivity, specificity and area under the receiver operating characteristic curves (AUROCs). Urinary glucose was not included in this analysis due to the high AUROCs. The optimal cut-off for calculating sensitivity and specificity was determined as the minimum distance to the top left-hand corner [21]. Significance was set at $P < 0.05$. IBM SPSS version 21 statistical software (IBM Corp) was used to perform the logistic regression and ROC analyses.

Metabolic phenotypes by k-means algorithm

Cluster analysis to identify metabolic phenotypes, or metabo-types, was performed using the k-means cluster algorithm in MetaboAnalyst 3.0 [12,22]. This generated two clusters in the diabetes patients and two clusters in the non-diabetic participants by taking as inputs the identified metabolites from the OSC–PLS-DA analysis and applying the k-means clustering algorithm [12]. After k-means analysis, the results for the four clusters were visualized using hierarchical clustering analysis.

Results

Subjects' characteristics

Our participants were 67 ± 6 years old and nearly one-third were male (Table 1). Also, 55% of participants had T2D and 47% were obese. They were divided according to T2D diagnosis, as previously reported [14,16]. Both groups (T2D and non-T2D) were well balanced in terms of demographic characteristics and other cardiovascular risk factors, such as blood pressure, plasma

lipids, and antihypertensive and hypolipidaemic medications ($P > 0.05$). Otherwise, measures of waist circumference, plasma glucose and use of antidiabetic agents were significantly higher in the T2D patients, as expected.

Profiles of discriminant metabolites of T2D biomarkers by 1H-NMR metabolomics

OSC–PLS-DA models were applied to determine the profile of discriminant metabolites in T2D vs non-T2D subjects. These models resulted in one latent component with R2Y(cum) and Q2Y(cum) mean values of 0.829 and 0.679, respectively, indicating a good ability to classify individuals according to their T2D status. A permutation test ($n = 200$), with intercept R2Y and Q2Y mean values of 0.306 and -0.149, respectively, confirmed the validity of the model (Fig. S1; see supplementary materials associated with this article online). In addition, sensitivity, specificity and accuracy values were calculated from the OSC–PLS-DA models when samples were predicted ($n = 5$); these values were then included in a misclassification table (Table S1; see supplementary materials associated with this article online). Thus, t-test analyses among VIP > 1.0 identified 33 metabolites that were significantly different between the T2D and non-T2D participants (Table 2). Of these metabolites, 17 were significantly increased in T2D patients compared with non-T2D subjects, while the remaining 16 metabolites were decreased in T2D patients. In addition, the metabolic fingerprint associated with T2D was found to be significantly independent of waist circumference except for 4-deoxythreonic acid and citrate ($P = 0.11$), and 3- hydroxybutyrate (3HB) ($P = 0.062$; Table 2). Furthermore, no statistical differences were observed in levels of metabolites among T2D patients whether taking drug treatment or not (data not shown). A comprehensive analyses of the metabolic pathways (P and impact values) revealed that the carbohydrate and amino-acid pathways were the most altered among T2D patients (Fig. S2, Table S2; see supplementary materials associated with this article online). The metabolites involved in these pathways can be up- and downregulated (Fig. S3; see supplementary materials associated with this article online), and each metabolite is related to its own pathway (Table S3; see supplementary materials associated with this article online).

Multi-metabolite signature of T2D prevalence

The multi-metabolite signature for better discrimination of T2D prevalence included higher levels of methylsuccinate, alanine, dimethylglycine and guanidinoacetate, as well as lower levels of glutamine, methylguanidine, 3-hydroxymandelate and hippurate (Table S4; see supplementary materials associated with this article online). In the validation set, the specificity and sensitivity of the multi-metabolite signature were 87.0% and 96.4%, respectively, while the AUROC was 96.4% (95% CI: 92.0–100%; $P < 0.001$). However, the specificity, sensitivity and AUROC values of each individual metabolite as well as urinary glucose were lower than those of the multi-metabolite signature (Fig. 1, Table 3).

Characterization of metabotypes

Unsupervised analysis of k-means gave two metabotypes of diabetes participants and two metabotypes of non-diabetes participants (Table S5; see supplementary materials associated with this article online) from data for the 33 identified metabolites. After determining those four metabotypes, the results were visualized using hierarchical clustering (heatmap) analysis (Fig. 2), where samples/individuals are shown on the x-axis and metabolites are displayed on the y axis. Most of the up-and downregulated metabolites observed in the clusters were similar to those reported in Table 2 for T2D and non-T2D participants except for four metabolites: acetoacetate (AA); p-cresol; phenylalanine; and phenylacetylglutamine (PAG). Levels of these four metabolites were significantly higher in clusters 2 and 3 than in clusters 1 and 4 (Fig. 2; $P < 0.05$) on stratifying the entire cohort, and were orthogonal for T2D. Thus, the two metabotypes of T2D (clusters 1 and 2) and two metabotypes of non-T2D (clusters 3 and 4) differed in these four metabolites. Cluster differences for subjects' characteristics, concentrations of biochemical parameters and use of medication are presented in Table S5 (see supplementary materials associated with this article online). The main difference was that cluster 2, followed by cluster 1, had the highest plasma glucose levels, and both were significantly different from clusters 3 and 4 ($P < 0.001$). As expected, the use of insulin and oral antidiabetic agents was significantly different between T2D and non-T2D participants ($P < 0.001$), but did not differ between T2D metabotypes ($P = 0.20$).

Discussion

The present study found significant differences in the profile of 33 urinary metabolites between T2D and non-T2D participants, using a 1H-NMR-based, non-targeted metabolomic approach. Specifically, a model of eight metabolites was the multi-metabolite signature that discriminated between T2D and non-T2D after stepwise logistic regression analysis and AUROC evaluation. To the best of our knowledge, this was the first-ever study to use spot urine to determine the pathways altered in T2D in a free-living population, along with identifying a multi-metabolite signature of T2D prevalence while highlighting the key implied metabolites. This metabolomic clinical study also confirms the associated perturbations of amino-acid metabolism, with some amino acids being used as substrates for gluconeogenesis. In addition, the increased excretion of amino acids could indicate an increase in protein degradation [4]. This was observed in our present study, and corroborates other metabolomic studies showing enhanced excretion of the glucogenic amino-acids alanine [5,6] and phenylalanine [6] and derived metabolites such as guanidinoace-tate, and the decreased excretion of glutamine [6] and histidine [5]. Previously, it was found that levels of phenylalanine and glutamine were positively and inversely, respectively, associated with the risk of prediabetes and T2D [7,23]. In addition, deregulation of branched-chain amino-acid metabolites (valine, leucine, isoleucine) has also been associated with risk of diabetes and insulin resistance [7,23]. Indeed, such changes have been observed in urine through the increased excretion of metabolites such as 3-hydroxyisovalerate [24] and methylsuccinate from their degradation pathways, which may reflect greater isoleucine catabolism [25]. The present metabolomic clinical study has shown an increase in the glycolysis and gluconeogenesis pathways in the liver associated with increased excretion of metabolites, including lactate, glucose and pyruvate, as also observed in previous studies [4]. Increased amounts of some carboxylic acids, such as cis-aconitate, an intermediary in the tricarboxylic acid cycle, and dicarboxylic suberic acid, were observed in the urine of T2D patients. In fact, increased excretion of cis-aconitate reflects systemic stress caused by hyperglycaemia or local effects on tubular transport in the kidneys [5]. Metabolites related to methylamine metabolism, such as dimethylglycine and trime-thylamine N-oxide, are systemic breakdown products of choline [5] that, due to their osmoregulatory properties, may be linked to a hyperosmotic effect of glucose or indicate renal papillary dysfunction when found in high concentrations [26]. In our study, decreased urinary levels

of creatinine and its metabolite methylguanidine [27] were also observed, which could be related to alterations of glomerular filtration rate (GFR) in T2D with a possible decrease of muscle mass [28], although our participants were free of nephropathy. However, in a recent report, lower creatinine excretion rates were associated with all-cause mortality in diabetes patients and in nephropathy [28]. Diabetes and obesity are lifestyle-related disorders that could cause an increased incidence of gut dysbiosis [29], which is directly related to alterations in gut microbial-related metabolites [30]. Indeed, the results of our study have shown a reduction in the excretion of well-known microbial metabolites, such as hippurate, PAG and p-cresol [29], as well as trigonelline and 3-hydroxymandelate. While PAG and p-cresol are related to protein putrefaction, hippurate is a breakdown product of polyphenol and fibre metabolism [9], and 3-hydroxymandelate is a metabolite of tyrosine [19]. Previous studies had observed that individuals with impaired glucose tolerance and patients with T2D have lower levels of hippurate and PAG [5]. Therefore, our study supports previous findings that a microbiota imbalance could be key in the pathogenesis of a diabetic state and that healthy diets and/or lifestyle patterns directed towards improving microbiota quality are essential for preventing advanced pathological states [31]. The present study identified two distinct metabotypes in T2D patients (clusters 1 and 2) and two in the non-T2D participants (clusters 3 and 4) using k-means cluster analysis based on their identified metabolic profiles. It should be noted that the metabotype comprising higher levels of four metabolites (phenylalanine, PAG, p-cresol and AA) was found in the entire study population and was orthogonal for T2D. In particular, differences were observed in some parameters between clusters 1 and 2 (T2D patients) whereas no differences were noted between clusters 3 and 4 (non-T2D subjects). Although the increase in these metabolites were orthogonal for T2D, when the focus was on diabetes patients, those with higher levels of those four metabolites also had higher levels of plasma glucose, but with no differences in use of antidiabetic medications or in other characteristics. Thus, our hypothesis is that the T2D patients in cluster 2 could have had a greater lack of control over their disease which, in the long term, could have led to a greater number of complications such as myocardial infarction, stroke, heart failure and kidney disease [32]. Certainly, phenylalanine has been described as a marker of higher diabetes risk [7,23] and, furthermore, has also been used together with tyrosine and isoleucine to predict long-term future

297 cardiovascular events, an increased disposition towards atheroscle- rosis and perhaps even inducible
298 myocardial ischaemia [33]. In addition, phenylalanine has been identified as a biomarker associated
299 with future cardiovascular events in meta-analyses [34]. PAG and p-cresol are metabolites of
300 microbial origin [35]. PAG comes from the conversion of phenylalanine to phenylacetate by
301 microbiota and its subsequent conjugation with glutamine [36]; and p-cresol, the most widely studied
302 uraemic retention solute, is formed by microbial metabolism of tyrosine [36]. PAG has been
303 described as a strong independent risk factor for mortality and CVD in patients with chronic kidney
304 disease [35], while p-cresol has been described as a predictor of cardiovascular events independent
305 of GFR in patients with mild-to-moderate kidney disease [37]. The fourth metabolite that differed
306 between T2D clusters was the ketone body AA. This is generated from the ketogenic amino-acid
307 lysine and may also be derived from b-oxidation of fatty acids. AA and 3HB are at a ratio of 1:1 in a
308 physiological state, although 3HB increases its excretion in ketoacidosis [38]. Recent evidence has
309 highlighted the association between elevated levels of ketone bodies and hyperglycaemia and T2D
310 [39]. It is also worth noting that the T2D patients in clusters 1 and 2 had similar mean ratios of
311 AA:3HB (1:2), whereas clusters 3 and 4 (non-T2D) had mean ratios of 1:1 (albeit not statistically
312 significant). However, there were statistically significant differences ($P = 0.007$) between ratios in
313 T2D (1:2) vs non- T2D (1:1) participants. Both hyperketonaemia and ketosis have been related to
314 liver, brain and microvasculature complications, which can increase the risk of morbidity and
315 mortality [40]. Therefore, the subjects in clusters 2 and 3 with increased levels of these four
316 metabolites could have higher risks of CVD and other such events in future. Thus, further studies
317 should now evaluate these metabolites in such populations in long-term studies. One limitation of our
318 present study is that the panel of metabolites and the model used for the multi-metabolite signature
319 imprinting of T2D were obtained from a high-cardiovascular-risk population, and so needs to be
320 validated and replicated in other populations. In addition, the metabolite panel should also be tested
321 in patients with different grades of T2D, including prediabetes states, to determine its limit values for
322 prediction. Moreover, it would be of interest to evaluate whether our metabotypes are modified in
323 states such as prediabetes. Another limitation of our study is that the microbial composition in these
324 participants was unknown, thereby preventing any correlations with the identified metabolites. On

the other hand, one strength of our study is that it reproduced of real-life conditions of the participants. In conclusion, the results of our cross-sectional study using a non-targeted 1H-NMR metabolomics approach reveal a multi-metabolite signature of T2D prevalence comprising eight metab-olites belonging to pathways related mainly to glucogenic and ketogenic amino acids, glycolysis and gluconeogenesis, carboxylic acid metabolism and changes in gut microbiota metabolism. This is also the first study to identify metabotypes in T2D, revealing that such patients have higher levels of phenylalanine, PAG, p-cresol and AA—metabolites related to higher risks of long-term cardio-vascular events—and also higher levels of plasma glucose. Nevertheless, as they were orthogonal for T2D, further studies now need to evaluate their long-term effects. In addition, this study reinforces the use of metabolomics to discover and to evaluate the main metabolic pathways altered in T2D and the metabotypes of individuals. Thus, it would be highly useful to investigate T2D diagnosis and treatment to further support the development of stratified and precision medicine.

Authors' contributions

M.U.-S., R.L., R.E., D.C., J.V.S., J.S.-S and C.A.-L. conceived and designed the study; M.U.-S., E.A.-A., R.L. and R.V.-F. performed the analyses; M.U.-S., E.A., R.L., R.V.-F., F.C., C.A.-L. and A.S.-P. analyzed the results; M.U.-S. and E.A. drafted the article. All of the authors critically revised the manuscript for important intellectual content. M.U.-S. and C.A.-L. are the guarantors of this work and, as such, had full access to all of the study data and take responsibility for the integrity of the data and accuracy of the data analysis.

Disclosure of interest

The authors declare that they have no competing interest.

Acknowledgments

This research was supported by Spanish national grants from the Ministry of Economy and Competitiveness (MINECO), and co-funded by FEDER (Fondo Europeo de Desarrollo Regional): AGL2009-13906-C02-01, as well as RYC-2011-09677, Associacio´ Catalana de Diabetis (ACD2015), CIBERFES, CIBER 06/03, CNIC-06, PI13/01172 Project (Plan Nacional de I+D+i

2013–2016), PI11/ 02505 by ISCII-Subdireccio´n General de Evaluacio´n y Fomento de la Investigacio´ n, Fundacio la Marato TV3 (ref: 201608.10), and PROMETEO017/2017 (Generalitat Valenciana). We also thank the EU Joint Programming Initiative 'A Healthy Diet for a Healthy Life' on Biomarkers BioNHFOODBALL (PCIN-2014-133-MINECO-Spain) and the 2014SGR1566 award from the Generalitat de Catalunya's Agency AGAUR. M.U.-S. would like to thank the Ramo´n y Cajal programme from MINECO and Fondo Social Europeo. E.A.-A. would like to thank CONACYT (Mexico) for the PhD fellowship.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.diabet.2018. 02.006.

References

[1] American Diabetes Association. Classification and diagnosis of diabetes. Dia-betes Care 2017;40:S11–24.

[2] Paneni F, Beckman JA, Creager MA, Cosentino F. Diabetes and vascular disease: pathophysiology, clinical consequences, and medical therapy: part I. Eur Heart J 2013;34:2436–43.

[3] Adamski J. Key elements of metabolomics in the study of biomarkers of diabetes. Diabetologia 2016;59:2497–502.

[4] Urpi-Sarda M, Almanza-Aguilera E, Tulipani S, Tinahones F, Salas-Salvado J, Andres-Lacueva C. Metabolomics for biomarkers of Type 2 Diabetes Mellitus: advances and nutritional intervention trends. Curr Cardiovasc Risk Rep 2015;9:12.

[5] Salek RM, Maguire ML, Bentley E, Rubtsov DV, Hough T, Cheeseman M, et al. A metabolomic comparison of urinary changes in type 2 diabetes in mouse, rat, and human. Physiol Genom 2007;29:99–108.

[6] van Doorn M, Vogels J, Tas A, van Hoogdalem EJ, Burggraaf J, Cohen A, et al. Evaluation of metabolite profiles as biomarkers for the pharmacological effects of thiazolidinediones in Type 2 diabetes mellitus patients and healthy volun-teers. Br J Clin Pharm 2007;63:562–74.

[7] Herder C, Kowall B, Tabak AG, Rathmann W. The potential of novel biomarkers to improve risk prediction of type 2 diabetes. Diabetologia 2014;57:16–29.

[8] Pearson ER. Personalized medicine in diabetes: the role of 'omics' and bio-markers. Diabet Med 2016;33:712–7.

[9] Heinzmann SS, Merrifield CA, Rezzi S, Kochhar S, Lindon JC, Holmes E, et al. Stability and robustness of human metabolic phenotypes in response to sequential food challenges. J Proteome Res 2012;11(2):643–55.

[10] Cuperlovic-Culf M, Belacel N, Culf A, Chute I, Ouellette R, Burton I, et al. NMR metabolic analysis of samples using fuzzy k-means clustering. Magn Reson Chem 2009;47(Suppl 1):S96–104.

[11] Nicholson JK, Holmes E, Kinross JM, Darzi AW, Takats Z, Lindon JC. Metabolic phenotyping in clinical and surgical environments. Nature 2012; 491:384–92.

[12] Vazquez-Fresno R, Llorach R, Perera A, Mandal R, Feliz M, Tinahones FJ, et al. Clinical phenotype clustering in cardiovascular risk patients for the identifi-cation of responsive metabotypes after red wine polyphenol intake. J Nutr Biochem 2016;28:114–20.

[13] O'Donovan C, Walsh M, Nugent A, McNulty B, Walton J, Flynn A, et al. Use of metabotyping for the delivery of personalised nutrition. Mol Nutr Food Res 2015;59:377–85.

[14] Estruch R, Ros E, Salas-Salvado J, Covas MI, Corella D, Aros F, et al. Primary prevention of cardiovascular disease with a Mediterranean diet. N Eng J Med 2013;368:1279–90.

[15] Estruch R, Martinez-Gonzalez MA, Corella D, Salas-Salvado J, Ruiz-Gutierrez V, Covas MI, et al. Effects of a Mediterranean-style diet on cardiovascular risk factors: a randomized trial. Ann Intern Med 2006;145:1–11.

[16] Salas-Salvado J, Bullo M, Estruch R, Ros E, Covas MI, Ibarrola-Jurado N, et al. Prevention of diabetes with Mediterranean diets: a subgroup analysis of a randomized trial. Ann Intern Med 2014;160:1–10.

[17] Vazquez-Fresno R, Llorach R, Urpi-Sarda M, Lupianez-Barbero A, Estruch R, Corella D, et al. Metabolomic pattern analysis after Mediterranean Diet Intervention in a Nondiabetic Population: a 1- and 3-year follow-up in the PREDIMED study. J Proteome Res 2014;14:531–40.

[18] Xia J, Mandal R, Sinelnikov IV, Broadhurst D, Wishart DS. MetaboAnalyst 2.0 – a comprehensive server for metabolomic data analysis. Nucleic Acids Res 2012;40(Web Server issue):W127–33.

[19] Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0 – the human metabolome database in 2013. Nucleic Acids Res 2013;41(Database issue):D801–7.

[20] Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, et al. BioMagResBank. Nucleic Acids Res 2008;36(Database issue):D402–8.

[21] Garcia-Aloy M, Llorach R, Urpi-Sarda M, Tulipani S, Estruch R, Martinez- Gonzalez MA, et al. Novel multimetabolite prediction of walnut consumption by a urinary biomarker model in a free-living population: the PREDIMED study. J Proteome Res 2014;13:3476–83.

[22] Gawlik A, Shmoish M, Hartmann MF, Malecka-Tendera E, Wudy SA, Hochberg Z. Steroid metabolomic disease signature of nonsyndromic childhood obesity. J Clin Endocrinol Metab 2016;101:4329–37.

[23] Guasch-Ferre M, Hruby A, Toledo E, Clish CB, Martinez-Gonzalez MA, Salas- Salvado J, et al. Metabolomics in prediabetes and diabetes: a systematic review and meta-analysis. Diabetes Care 2016;39:833–46.

[24] Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res 2014;42(Database issue):D199–205.

[25] Nowaczyk MJ, Lehotay DC, Platt BA, Fisher L, Tan R, Phillips H, et al. Ethylma-lonic and methylsuccinic aciduria in ethylmalonic encephalopathy arise from abnormal isoleucine metabolism. Metabolism 1998;47:836–9.

[26] Guan M, Xie L, Diao C, Wang N, Hu W, Zheng Y, et al. Systemic perturbations of key metabolites in diabetic rats during the evolution of diabetes studied by urine metabonomics. PLoS ONE 2013;8(4):e60409.

[27] Ienaga K, Hum Park C, Yokozawa T. Daily hydroxyl radical scavenging capacity of mammals. Drug Discov Ther 2014;8(2):71–5.

[28] Sinkeler SJ, Kwakernaak AJ, Bakker SJ, Shahinfar S, Esmatjes E, de Zeeuw D, et al. Creatinine excretion rate and mortality in type 2 diabetes and nephrop-athy. Diabetes care 2013;36(6):1489–94.

[29] Nicholson JK, Holmes E, Kinross J, Burcelin R, Gibson G, Jia W, et al. Host-gut microbiota metabolic interactions. Science 2012;336(6086):1262–7.

[30] Herrema HRGIJ, Nieuwdorp M. Emerging role of intestinal microbiota and microbial metabolites in metabolic control. Diabetologia 2017;60:613–7.

[31] Blandino G, Inturri R, Lazzara F, Di Rosa M, Malaguarnera L. Impact of gut microbiota on diabetes mellitus. Diabetes Metab 2016;42:303–15.

[32] Monnier L, Colette C. Postprandial and basal hyperglycaemia in type 2 diabe-tes: contributions to overall glucose exposure and diabetic complications. Diabetes Metab 2015;41(6 Suppl 1). 6S9–6S15.

[33] Magnusson M, Lewis GD, Ericson U, Orho-Melander M, Hedblad B, Engstrom G, et al. A diabetes-predictive amino acid score and future cardiovascular disease. Eur Heart J 2013;34:1982–9.

[34] Wurtz P, Havulinna AS, Soininen P, Tynkkynen T, Prieto-Merino D, Tillin T, et al. Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. Circulation 2015;131:774–85.

[35] Poesen R, Claes K, Evenepoel P, de Loor H, Augustijns P, Kuypers D, et al. Microbiota-derived phenylacetylglutamine associates with overall mortality and cardiovascular disease in patients with CKD. J Am Soc Nephrol 2016; 27:3479–87.

[36] Swann JR, Spagou K, Lewis M, Nicholson JK, Glei DA, Seeman TE, et al. Microbial-mammalian cometabolites dominate the age-associated urinary metabolic phenotype in Taiwanese and American populations. J Proteome Res 2013;12:3166–80.

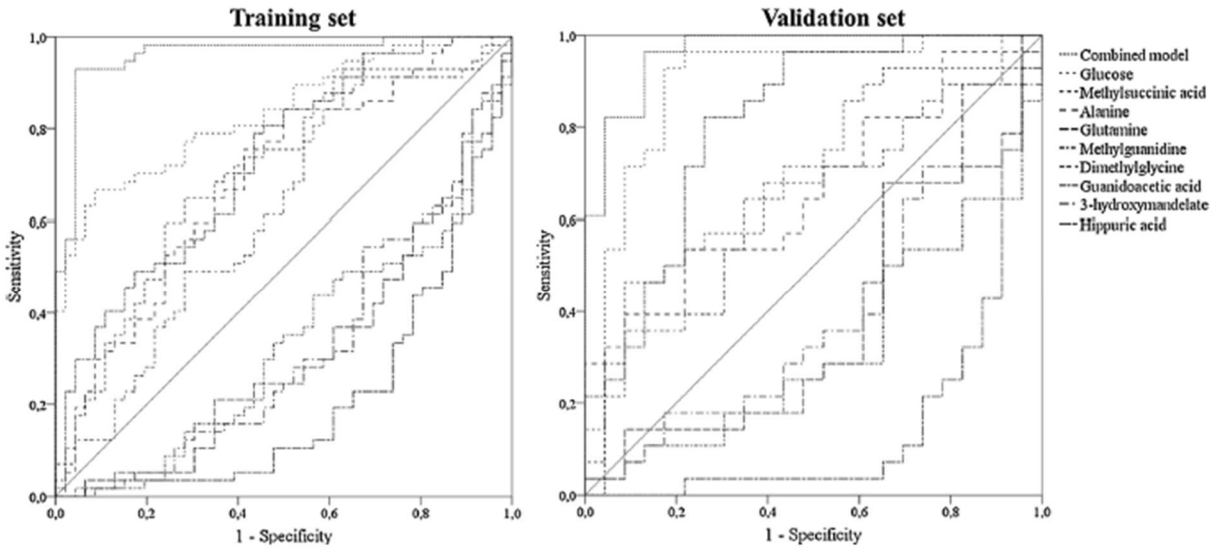
[37] Meijers BK, Claes K, Bammens B, de Loor H, Viaene L, Verbeke K, et al. p-Cresol and cardiovascular risk in mild-to-moderate kidney disease. Clin J Am Soc Nephrol 2010;5:1182–9.

[38] Sheikh-Ali M, Karon BS, Basu A, Kudva YC, Muller LA, Xu J, et al. Can serum beta-hydroxybutyrate be used to diagnose diabetic ketoacidosis? Diabetes care 2008;31:643–7.

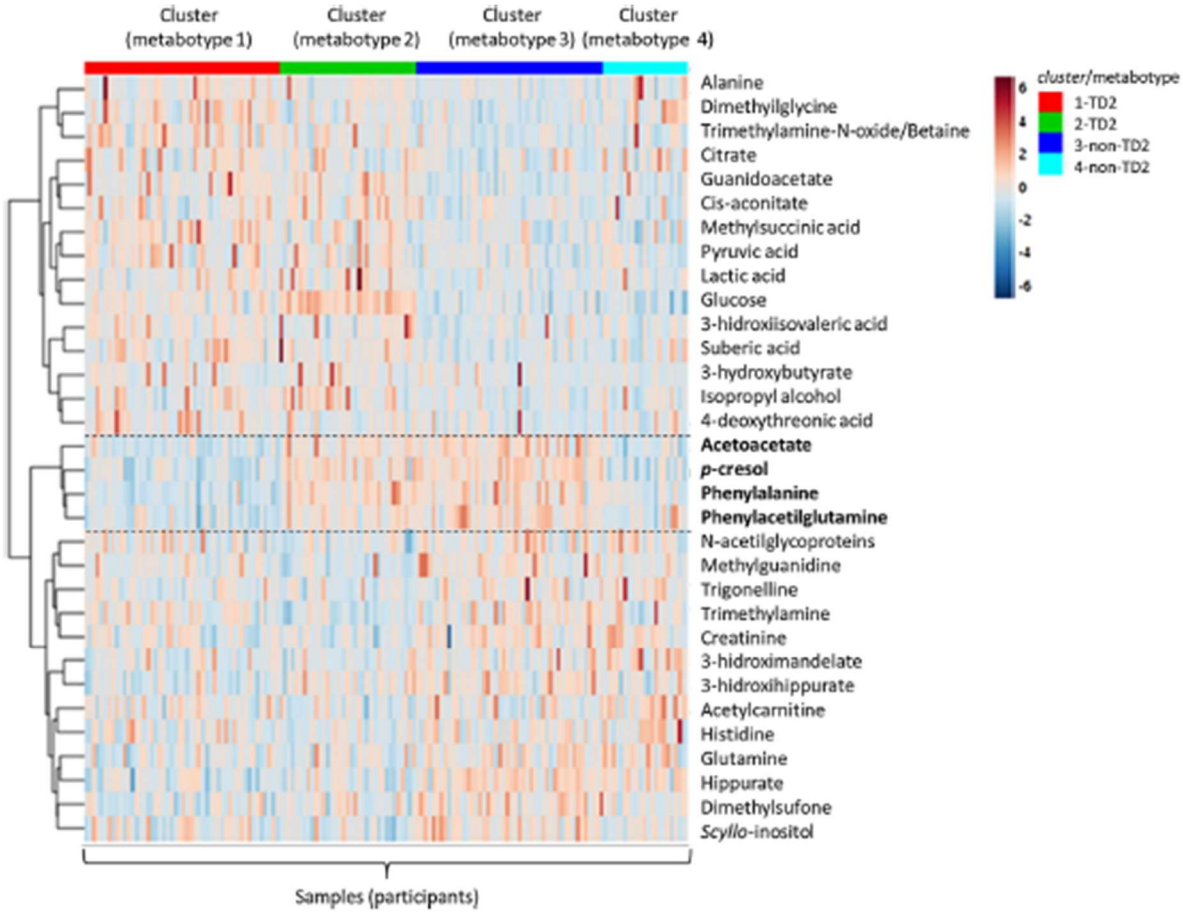
[39] Mahendran Y, Vangipurapu J, Cederberg H, Stancakova A, Pihlajamaki J, Soininen P, et al. Association of ketone body levels with hyperglycemia and type 2 diabetes in 9,398 Finnish men. Diabetes 2013;62:3618–26.

[40] Kanikarla-Marie P, Jain SK. Hyperketonemia and ketosis increase the risk of complications in type 1 diabetes. Free Radic Biol Med 2016;95:268–77.

468 **FIGURES**



469 **Fig. 1.** Area under the receiver operating characteristic curves (AUROCs) comparing the multi-metabolite signature and individual metabolites of T2D prevalence.



470 **Fig. 2.** Heatmap of metabotypes of participants in the Prevention with Mediterranean Diet (PREDIMED) study after *k*-means (cluster) analysis. *Abbreviations*: T2D, diabetic participant; non-T2D, non-diabetic participant.

471

TABLES

Table 1
Characteristics of the 154 participants according to type 2 diabetes (T2D) prevalence status.

	Total participants (n = 154)	T2D (n = 85)	Non-T2D (n = 69)	P
Age (years)	67 ± 5.7	67 ± 6.0	67 ± 5.5	0.67
Men, n (%)	48 (31)	30 (35)	18 (22)	0.15
Current smokers, n (%)	20 (13)	11 (13)	9 (13)	0.58
Body mass index (BMI), kg/m^2	30.18 ± 4.29	30.40 ± 4.60	30.0 ± 3.92	0.50
Obesity (BMI ≥ 25 kg/m^2), n (%)	71 (47)	37 (44)	34 (49)	0.29
Weight, kg	76.0 ± 12.67	77.71 ± 13.27	73.90 ± 11.65	0.063
Waist circumference, cm	101.31 ± 11.38	103.71 ± 11.34	98.35 ± 10.81	0.003
Physical activity, MET min/day	261.27 ± 268.32	241.22 ± 285.51	285.96 ± 245.31	0.30
Systolic blood pressure, mmHg	144.01 ± 19.72	140.70 ± 17.97	148.0 ± 21.20	0.08
Diastolic blood pressure, mmHg	81.94 ± 10.22	81.06 ± 10.46	83.0 ± 9.95	0.38
Mediterranean diet score	8.5 ± 1.8	8.383 ± 1.8	8.7 ± 1.8	0.19
Plasma biomarkers, mg/dL				
Glucose	128.2 ± 48.5	159.3 ± 45.3	89.9 ± 7.2	<0.001
HDL cholesterol	55.1 ± 14.4	51.5 ± 14.4	58.4 ± 13.9	0.09
LDL cholesterol	131.7 ± 28.8	123.3 ± 29.7	139.0 ± 26.5	0.06
Total cholesterol	213.6 ± 14.4	206.9 ± 35.0	218.8 ± 32.9	0.21
Triglycerides	150.0 ± 113.6	176.6 ± 153.5	125.8 ± 49.3	0.13
Medications, n (%)				
Antihypertensive agents	65 (42)	32 (38)	33 (48)	0.13
Hypolipidaemic agents	85 (55)	42 (49)	43 (62)	0.07
Oral antidiabetic agents	65 (42)	65 (76)	0	<0.01
Insulin	18 (12)	18 (21)	0	<0.01

Data are means ± SD for continuous variables and n (%) for categorical variables.
MET: metabolic equivalent; HDL/LDL: high-density/low-density lipoprotein.

Table 2
Urinary metabolites identified by ^{1}H-NMR in type 2 diabetes (T2D) and non-T2D participants in the PREDIMED study.

Metabolite[*]	Multiplicity (ppm) and *J* coupling (Hz)[d]	Excretion: T2D *vs* non-T2D	*P*[b]	*P*[c]	VIP	VIP-sd(VIP)[a]
Higher in T2D participants						
3-Hydroxyhippuric acid	7.12 (d, *J* = 8.92), 7.29 (m), 7.37 (m), 7.41 (m)	↑	0.011	0.017	1.1	0.9
3-Hydroxybutyrate	1.19 (d, *J* = 6.24)	↑	0.047	0.062	0.9	1.2
3-Hydroxyisovalerate	1.27 (s)	↑	0.004	0.011	1.3	1.0
4-Deoxythreonic acid	1.22 (d, *J* = 6.47)	↑	0.039	0.11	1.0	1.0
Alanine	1.48 (d, *J* = 7.27)	↑	0.013	0.026	1.5	0.9
cis-Aconitate	3.13 (s), 5.75 (s)	↑	0.004	0.010	1.4	1.4
Citrate	2.54 (d, *J* = 15.16), 2.68 (d, *J* = 14.67)	↑	0.047	0.11	1.3	1.8
Dimethylglicine	2.93 (s)	↑	0.014	0.040	1.5	1.2
Glucose	3.25 (m), 3.45 (m), 3.50 (m), 3.72 (m), 4.64 (d, *J* = 7.93), 5.25 (d, *J* = 3.71)	↑	<0.001	<0.001	3.6	3.5
Guanidinoacetate	3.79 (s)	↑	<0.001	<0.001	2.8	2.8
Isopropyl alcohol	1.16 (d, *J* = 6.24)	↑	0.002	0.007	1.3	1.2
Lactate	1.33 (d, *J* = 6.90), 4.11 (q)	↑	<0.001	0.001	2.0	1.6
Methylsuccinate	1.07 (d, *J* = 7.01)	↑	<0.001	<0.001	1.8	1.8
Phenylalanine	7.32 (d, *J* = 7.15), 7.36 (m), 7.43 (t)	↑	0.009	0.032	1.2	1.2
Pyruvate	2.38 (s)	↑	<0.001	<0.001	1.6	1.7
Suberic acid	1.29 (m), 1.59 (t)	↑	<0.001	<0.001	1.8	1.5
TMAO/betaine[e]	3.27 (s), 3.90 (s)	↑	<0.001	<0.001	2.7	1.8
Lower in T2D participants						
3-Hydroxymandelate	6.87 (d, *J* = 8.27), 6.93 (t), 6.99 (d, *J* = 7.64)	↓	0.003	0.004	1.7	1.1
Acetoacetate	2.28 (s)	↓	0.007	0.007	1.1	1.1
Acetylcartinine	2.15 (s), 2.19 (s)	↓	0.015	0.010	1.3	1.1
Creatinine	4.06 (s)	↓	<0.001	<0.001	2.4	2.1
Dimethylsulphone	3.16 (s)	↓	<0.001	<0.001	1.9	1.9
Glutamine	2.14 (m), 2.47 (m)	↓	0.001	0.002	1.3	1.2
Hippurate	3.98 (d, *J* = 5.82) 7.55 (t) 7.63 (tt) 7.83 (dd) 8.54 (bb)	↓	<0.001	<0.001	2.8	2.4
Histidine	7.08 (s)	↓	0.011	0.009	1.4	0.8
Methylguanidine	2.82 (s)	↓	0.001	<0.001	1.7	1.6
N-acetylglutamine	2.09 (m), 4.16 (m), 2.25 (t), 7.36 (m), 7.98 (m)	↓	0.004	0.008	1.1	0.9
N-acetylglycoproteins	2.04 (s)	↓	0.017	0.043	1.1	1.0
p-Cresol	2.34 (s), 7.22 (d, *J* = 8.37), 7.27 (d, *J* = 8.90)	↓	0.009	0.004	1.7	1.2
Phenylacetylglutamine	2.27 (t), 4.17 (m), 7.42 (m), 7.26 (m), 7.98 (bb)	↓	0.026	0.025	1.1	1.2
Scyllo-inositol	3.36 (s)	↓	<0.001	0.001	2.1	1.9
Trigonelline	4.43 (s) 8.08 (m) 8.84 (m) 9.12 (s)	↓	0.002	0.004	1.4	1.3
TMA	2.86 (s)	↓	0.005	0.005	1.3	0.5

[*] Identified by variable importance in projection (VIP) values ≥1.0 on orthogonal signal correction with partial least squares discriminant analysis (OSC-PLS-DA) model.
[a] VIP minus standard deviation of VIP (used as additional quality parameter)
[b] Adjusted by Benjamini–Hochberg procedure for multiple comparisons.
[c] Adjusted by waist circumference.
[d] s: singlet, d: doublet, t: triplet, q: quadruplet, m: multiplet, bb: broad band.
[e] Overlapping peaks.
PREDIMED: Prevention with Mediterranean Diet; TMAO: Trimethylamine N-oxide; TMA: Trimethylamine.

Table 3
Receiver operating characteristic (ROC) curve parameters for the prediction model and individual metabolites.

	AUC (95% CI)	*P*	Specificity (%)	Sensitivity (%)
Training set				
Combined model	96.1 (92.3–100)	1.11×10^{-15}	96.0	93.0
Glucose	83.5 (76.0–91.1)	5.50×10^{-9}	91.3	66.7
Validation set				
Combined model	96.4 (92.0–100)	1.11×10^{-8}	87.0	96.4
Methylsuccinate	69.4 (54.8–84.0)	0.018	74.0	57.1
Alanine	66.3 (51.5–81.1)	0.047	78.3	53.6
Glutamine	37.6 (21.8–53.4)	0.130	34.8	67.8
Methylguanidine	31.8 (17.1–46.5)	0.027	34.8	50.0
Dimethylglycine	64.0 (48.7–79.3)	0.088	65.2	64.3
Guanidinoacetate	81.0 (68.8–93.4)	<0.001	74.0	82.1
3-Hydroxymandelate	40.2 (24.3–56.2)	0.233	30.0	64.3
Hippurate	15.0 (3.4–36.4)	<0.001	8.69	75.0
Glucose	89.8 (80.2–99.3)	1.26×10^{-6}	82.6	93.0

AUC: area under the curve.