European Journal of Clinical Nutrition (2013), 1–9

www.nature.com/ejcn

ORIGINAL ARTICLE

Variation of serum metabolites related to habitual diet: a targeted metabolomic approach in EPIC-Potsdam

A Floegel[1], A von Ruesten[1,2], D Drogan[1], MB Schulze[3], C Prehn[4], J Adamski[4], T Pischon[5] and H Boeing[1]

BACKGROUND/OBJECTIVE: Serum metabolites have been linked to higher risk of chronic diseases but determinants of serum metabolites are not clear. We aimed to investigate the association between habitual diet as a modifiable risk factor and relevant serum metabolites.

SUBJECTS/METHODS: This cross-sectional study comprised 2380 EPIC-Potsdam participants. Intake of 45 food groups was assessed by food frequency questionnaire and concentrations of 127 serum metabolites were measured by targeted metabolomics. Reduced rank regression was used to find dietary patterns that explain the maximum variation of metabolites.

RESULTS: In the multivariable-adjusted model, the proportion of explained variation by habitual diet was ranked as follows: acyl-alkyl-phosphatidylcholines (5.7%), sphingomyelins (5.1%), diacyl-phosphatidylcholines (4.4%), lyso-phosphatidylcholines (4.1%), acylcarnitines (3.5%), amino acids (2.2%) and hexose (1.6%). A pattern with high intake of butter and low intake of margarine was related to acylcarnitines, acyl-alkyl-phosphatidylcholines, lyso-phosphatidylcholines and hydroxy-sphingomyelins, particularly with saturated and monounsaturated fatty acid side chains. A pattern with high intake of red meat and fish and low intake of whole-grain bread and tea was related to hexose and phosphatidylcholines. A pattern consisting of high intake of potatoes, dairy products and cornflakes particularly explained methionine and branched chain amino acids. Dietary patterns related to type 2 diabetes-relevant metabolites included high intake of red meat and low intake of whole-grain bread, tea, coffee, cake and cookies, canned fruits and fish.

CONCLUSIONS: Dietary patterns characterized by intakes of red meat, whole-grain bread, tea and coffee were linked to relevant metabolites and could be potential targets for chronic disease prevention.

European Journal of Clinical Nutrition advance online publication, 14 August 2013; doi:10.1038/ejcn.2013.147

Keywords: metabolomics; metabolites; diet; food intake; reduced rank regression; systems epidemiology

INTRODUCTION

Advancement of technologies from analytical chemistry, particularly nuclear magnetic resonance spectroscopy and mass spectrometry (MS), made high-throughput metabolomic analysis of biological specimen possible. To date, an increasing number of metabolomic platforms enable robust and quick measurements of metabolites not only for single patients but also for large-scale metabolic profiling in epidemiologic studies. Thereby, metabolomics offers great potential to study metabolic alterations that are linked to higher risk of chronic diseases. Alterations in blood concentrations of metabolites were for example observed among obese individuals, people with impaired glucose tolerance and type 2 diabetes patients.[1–5] These changes included acylcarnitines,[1,3] amino acids,[3,4] sugars[2,5] and different lipid species.[2,3] Recently, the use of tandem MS techniques in the frame of different prospective cohort studies has enabled the identification of metabolites that predict chronic diseases. Particularly, higher concentrations of branched chain and aromatic amino acids were identified to be linked to higher risk of type 2 diabetes in the Framingham Offspring study and the Malmö Diet and Cancer study.[6] Recent work from our group confirmed this finding; additionally, we identified increased hexose and diacyl-phosphatidylcholines, and reduced acyl-alkyl- and lyso-phosphatidylcholines, and sphingomyelins to be associated with development of type 2 diabetes in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study.[7]

From a public health perspective it is of interest to investigate modifiable risk factors that contribute to changes in metabolite concentrations related to disease risk. In this context, it has been suggested that metabolites may be very responsive to exposure of dietary, environmental and lifestyle factors.[8] Particularly, one would expect that habitual dietary patterns are linked to blood concentrations of metabolites, as diet could be a primary source of metabolites, and secondarily, induce metabolic responses.[9,10] In fact, a previous study suggested that self-reported nutrition

[1]Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany; [2]Department of Public Health and Community Medicine, University of Gothenburg, Gothenburg, Sweden; [3]Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany; [4]Institute of Experimental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany and [5]Molecular Epidemiology Group, Max Delbrück Center for Molecular Medicine (MDC) Berlin-Buch, Berlin-Buch, Germany. Correspondence: Dr A Floegel, Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Arthur-Scheunert-Allee 114-116, 14558 Nuthetal, Germany.
E-mail: anna.floegel@dife.de

Contributors: AF designed the research, conducted the data analysis, interpreted the data, wrote the manuscript and had primary responsibility for data integrity and the final content; AvR, DD, MBS contributed to the conception of this study, helped with the interpretation of the results and provided critical comments on the manuscript; CP and JA conducted metabolomics measurements and provided critical comments on the manuscript; TP and HB obtained funding, designed the research, helped with the interpretation of the results and provided critical comments on the manuscript.
Received 5 April 2013; revised 11 July 2013; accepted 12 July 2013

habits may be reflected in metabolic profiles.[11] However, the association between habitual diet and blood concentrations of different classes of metabolites related to disease risk has not been well examined at the population level. It would be of particular interest to identify those foods in the usual diet that explain most of the variation of metabolites. Thereby, the understanding of the biological mechanisms could be improved, and consequently, adequate dietary recommendations and prevention strategies for chronic disease risk could be developed.

Thus, to investigate the association between habitual diet and serum metabolites, we applied a common method from dietary pattern analysis, that is, reduced rank regression (RRR),[12] and used data on self-reported diet and measurement of serum metabolite concentrations from the EPIC-Potsdam study. In a first step, we investigated the proportion of variation in relevant serum metabolites (including acylcarnitines, amino acids, hexose and choline-containing phospholipids) that was related to dietary patterns derived from RRR. Second, we aimed to identify important foods that were included into these metabolite-linked dietary patterns.

SUBJECTS AND METHODS

Study population

EPIC is a multicenter prospective cohort study conducted in 10 European countries to investigate the relationship between diet and risk of cancer but considering other risk factors and diseases as well. EPIC-Potsdam located in Germany comprises 27 548 participants from the general population of Potsdam and surrounding areas with an age range mainly between 35 and 65 years at the time of recruitment.[13] Between 1994 and 1998, participants were invited to the study center where they provided written informed consent before they took part in the baseline examination.[14] During the baseline examination, anthropometric and blood pressure measurements and a blood sample collection were conducted by qualified medical staff following standardized protocols.[14] In particular, 30 ml of venous blood were drawn, immediately processed and fractionated, and serum, plasma, 'buffy coat' (leukocytes) and erythrocytes were stored in 0.5 ml straws at $-196\,^{\circ}C$.[14] The participants further completed an interview on medical conditions, a sociodemographic and lifestyle questionnaire, and a self-administered validated food frequency questionnaire (FFQ). The EPIC-Potsdam study was approved by the ethics committee of the Medical Society of the State of Brandenburg.

Within EPIC-Potsdam, a subcohort for biomarker measurements was established by randomly selecting 2500 participants from all EPIC-Potsdam participants who had provided blood samples at baseline ($n = 26444$).[15] By randomly selecting the subcohort the results are expected to be representative for the full cohort without the need to measure the biomarkers in the full cohort. For the present analysis, participants of the subcohort with implausible energy intakes (<800 and >6000 kcal/day) or missing data on serum metabolite concentrations and covariates were excluded ($n = 120$). Participants included into the analysis were not markedly different from the original study sample.

Dietary assessment

Habitual diet of the participants was assessed by a self-administered, semiquantitative FFQ at baseline. The FFQ contained 148 items and inquired about frequency and portion size of food and beverage consumption during the preceding 12 months. The frequencies ranged from never to five times per day or more. The participants had to assign their individual portion sizes relative to standard portion sizes. To facilitate the estimation of the individual portion size, standard portion sizes and household measures were illustrated with photos for different foods and dishes. Additional questions of the FFQ inquired about the usual fat content of dairy products and fat quality for food preparation. Usual intake of each food item in grams per day was calculated as the product of frequency and portion size. The FFQ used in the German EPIC-centers has been validated and was shown to generate reproducible results.[16–19]

Measurement of serum metabolites

Concentrations of metabolites were measured in baseline serum samples of the EPIC-Potsdam subcohort with the Absolue*IDQ* p150 kit (BIOCRATES, Innsbruck, Austria) by flow injection analysis tandem mass spectrometry (FIA)-MS/MS.[20] The targeted metabolomics method simultaneously determined concentrations of 163 predefined metabolites including acylcarnitines (C*x*:*y*), amino acids, hexose (sum of six-carbon monosaccharides without distinction of isomers) and choline-containing phospholipids (lyso-phosphatidylcholines, diacyl- and acyl-alkyl-phosphatidylcholines and sphingomyelins). For the lipid derivates, fatty acid side chains were abbreviated C*x*:*y*, where *x* represented the cumulative number of carbon atoms and *y* the cumulative number of double bonds. All samples were analyzed at the Genome Analysis Center (Helmholtz Zentrum München) in 2010. Sample preparation and metabolite quantification of these cohort samples has been described previously.[7] In brief, a robotic system (Hamilton ML star, Bonaduz, Switzerland) conducted the following steps: first, pipetting of 10 μl of serum onto filters with stable isotope-labeled internal standards on a 96-well plate; second, drying of the plates in nitrogen stream; third, derivatization of the amino acids with 5% phenylisothiocyanat reagent; fourth, drying of the plates again; fifth, extraction of the remaining metabolites and internal standards using 5 mM ammonium acetate in methanol; sixth, centrifugation and filtration; and last, dilution of the final extracts with MS running solvent. Analysis of final extracts was performed with an API4000 triple quadrupole mass spectrometer (ABSciex, Framingham, MA, USA). Multiple reactions monitoring in combination with internal standards was applied for the quantification of metabolites, and concentrations were calculated in 'μM' using the Met*IQ* software package (BIOCRATES, Innsbruck, Austria).

The metabolomic method has been validated by the manufacturer in conformance with the FDA Guideline 'Guidance for Industry—Bioanalytical Method Validation, May 2001'[21], which implies proof of reproducibility within a given error range. Analytical specifications for limit of detection and evaluated quantification ranges, further limit of detection for semiquantitative measurements, identities of quantitative and semiquantitative metabolites, specificity, potential interferences, linearity, precision and accuracy, reproducibility and stability were described in Biocrates manual AS-P150. The limits of detection were set to three times the values of the 'zero samples' containing buffer. The lower and upper limit of quantification was determined experimentally by Biocrates. The analytical variation of the EPIC-Potsdam samples was additionally evaluated by measuring 230 replicates of one quality control sample, and the median coefficients of variation were 7.3% within-plate and 11.3% between-plates, respectively.[22] Those metabolites with very high analytical variation and below their limit of detection (mainly hydroxyacylcarnitines) were excluded from the present analysis; thus, the final metabolite set comprised 127 metabolites (17 acylcarnitines, 14 amino acids; 1 hexose; 34 diacyl-phosphatidylcholines; 37 acyl-alkyl-phosphatidylcholines; 10 lyso-phosphatidylcholines; and 14 sphingomyelins).

Statistical analysis

Descriptive statistics of the study population were obtained by calculating arithmetic mean and standard deviation for continuous variables or percentage for categorical variables. Habitual diet was summarized in 45 predefined food groups excluding alcoholic beverages, which have been previously used in the EPIC-Potsdam study,[23,24] and was reported as median and interquartile range. The correlation of responses was evaluated by calculating simple Spearman correlation coefficients for the individual metabolites of each class of metabolites and illustrated as circular network graphs with yEd graph editor (yWorks GmbH, Tuebingen. www.yworks.com). As most of the metabolites were normally distributed, they were not transformed. To evaluate the relationship between habitual diet and serum metabolites, a RRR was conducted with the PLS procedure implemented in SAS (statistical analysis software, version 9.2, SAS Institute Inc, Cary, NC, USA). RRR is a statistical method that is frequently adopted when combining dietary pattern analysis with biomarker data;[25,26] however, to our knowledge it has not yet been used for metabolomic research. Briefly, it determines linear functions of a set of predictors (such as foods) that explain a maximum of the variation in responses (for example, biomarkers and metabolites).[12] In our RRR model, intakes of 45 food groups (continuous in g/day) were considered as the predictor variables, and the metabolites (continuous in μmol/l) were used as response variables in an exploratory approach. Thereby, an

individual response set was defined for each class of metabolites (acylcarnitines $n=17$, amino acids $n=14$; hexose $n=1$; diacyl-phosphatidylcholines $n=34$; acyl-alkyl-phosphatidylcholines $n=37$; lyso-phosphatidylcholines $n=10$ and sphingomyelins $n=14$). Thus, dietary patterns that explain a maximum of the variation of the different classes of metabolites were derived. The total number of extracted diet factor scores equaled the number of individual metabolites, and therefore, depended on the class of metabolites. First, we looked at the variation of metabolites that was explained by all diet factor scores and compared the crude model with different multivariable-adjusted models. Data on food intake was adjusted by taking the residuals from linear regression analysis, entering the covariates age (years) and sex, and a multivariable model with additional adjustment for alcohol intake from beverages (non-consumers; women: >0–6 g/day, 6–12 g/day, >12 g/day; men: >0–12 g/day, 12–24 g/day, >24 g/day), smoking (never, former, current ⩽20 cigarettes/day, current >20 cigarettes/day), physical activity (average of cycling and sports during summer and winter season in hours/week), education (no degree/vocational training, trade/technical school and university degree), prevalent hypertension (*y*/*n*), prevalent diabetes (*y*/*n*), body mass index (kg/m^2) and waist circumference (cm). In the next step, the first two diet factor scores were explored in more detail, as they explained most of the variation of serum metabolites. Important foods were defined as those with absolute values of factor loadings ⩾|0.2|. The proportion of variation (%) of individual metabolites that was explained by the first two diet factor scores was retrieved. In a sensitivity analysis, the main analysis was repeated when stratified by sex. Next to this exploratory approach that focused on metabolite classes, we also used a hypothesis-driven approach. We repeated the analysis and included two metabolite factors as response variables that have previously been derived by principal component analysis and were linked to risk of incident type 2 diabetes in this population.[7] Thereby, we derived dietary patterns that may be particularly relevant for chronic disease risk.

RESULTS

The present study population comprised 2380 adults (61% women) with a mean age of ~50 years, mean body mass index of 26.1 kg/m^2 and mean concentrations of serum metabolites ranged from 45.5 μmol/l for acylcarnitines to 4828 μmol/l for hexose (Table 1). The median intake of 45 foods is presented in Supplementary Table S1 and the median concentrations of individual serum metabolites of this population are presented in Supplementary Table S2. Most of the metabolites of the same class were strongly positively correlated (Supplementary Figure S1).

For each class of metabolites, individual diet factor scores were calculated, which explained most of the variation of metabolites. Overall, the mean proportion of variation of metabolites that was explained by habitual diet ranged from 3.6% for amino acids to 7.7% for acyl-alkyl-phosphatidylcholines in the crude model (Figure 1). It was further decreased after adjustment of food intake for age and sex and was lowest after multivariable adjustment for other covariates. In the multivariable-adjusted model the mean explained variation of habitual diet was highest for acyl-alkyl-phosphatidylcholines, sphingomyelins, diacyl-phosphatidylcholines and lyso-phosphatidylcholines (5.7, 5.1, 4.4 and 4.1%; respectively). The classes of acylcarnitines, amino acids and hexoses were less explained by habitual diet (3.5, 2.2 and 1.6%; respectively). To best account for the potential confounders, only the multivariable-adjusted model was considered for further analysis. The major proportion of variation was explained by the first two diet factor scores (Figure 2).

The class of acylcarnitines was associated with a dietary pattern that included high intake of butter and low intake of margarine (Table 2). This pattern was particularly related to higher saturated acylcarnitines, such as C9, C16 and C18, and lower unsaturated acylcarnitines (C8:1, C18:2 and C14:2). A second pattern low in fish, vegetables and whole-grain bread was particularly linked to C14, C2 and C18 acylcarnitines (Table 3). A food pattern consisting of high intake of potatoes, cornflakes, dairy products, raw vegetables and desserts and low intake of soup and beverages explained the maximum variation of amino acids, and particularly methionine, besides branched chain and aromatic amino acids (Table 2). Serine was more strongly related to the second dietary pattern with higher intake of canned fruit, fried potatoes and legumes and lower intake of water, low-fat cheese and fish (Table 3). Hexose was particularly associated with a dietary pattern characterized by high intake of red meat, non-whole-grain bread and fish and low intake of whole-grain bread and tea (Table 2). A dietary pattern characterized by high intake of fish and poultry and low intake of sweet foods, margarine, tea and whole-grain bread was related to variation of diacyl-phosphatidylcholines (Table 2). This pattern best explained polyunsaturated and saturated diacyl-phosphatidylcholines C38:6, C36:0, C38:0 and C40:6. In contrast, a pattern with low intake of fish, whole-grain bread and tea, and higher intake of sauce and butter was associated with the mono-unsaturated diacyl-phosphatidylcholines C36:1, C28:1 and C34:1 (Table 3). The acyl-alkyl-phosphatidylcholines, particularly C34:0, C36:1, C34:1 and C30:0, were linked to a dietary pattern consisting of high intake of butter, red meat and high-fat dairy products and low intake of margarine and whole-grain bread (Table 2). The unsaturated acyl-alkyl-phosphatidylcholines were associated with a dietary pattern with higher intake of meats and margarine and lower intake of butter and sweet foods (Table 3). The class of lyso-phosphatidylcholines was related to a pattern characterized by high intake of butter, high-fat-dairy and sweet bread spreads and low intake of margarine, low-fat cheese and pasta/rice (Table 2). This dietary pattern was particularly related to

Table 1. Baseline characteristics of the EPIC-Potsdam subcohort ($n=2380$)

	Mean (s.d.) or %
Women (%)	61.1
Age (years)	49.8 (8.9)
University degree (%)	38.7
Physical activity[a] (hour/week)	2.8 (3.4)
Smoking (%)	
Never	47.0
Former	32.4
Current ⩽20 cigarettes/day	18.4
Current >20 cigarettes/day	2.2
Alcohol consumption[b] (%)	
Non-consumers	2.9
Low	45.5
Moderate	24.2
High	27.4
BMI (kg/m^2)	26.1 (4.3)
Waist circumference (cm)	
Men	94.1 (9.9)
Women	80.6 (11.7)
Prevalent hypertension (%)	48.6
Prevalent type 2 diabetes (%)	4.5
Serum metabolites (μmol/l)[c]	
Acylcarnitines ($n=17$)	45.5 (10.7)
Amino acids ($n=14$)	2323.9 (394.9)
Diacyl-phosphatidylcholines ($n=34$)	1880.2 (373.5)
Acyl-alkyl-phosphatidylcholines ($n=37$)	183.1 (36.2)
Lyso-phosphatidylcholines ($n=10$)	217.3 (49.9)
Sphingomyelins ($n=14$)	286.6 (61.6)
Hexose ($n=1$)	4828.0 (1352.5)

Abbreviation: BMI, body mass index. [a]Average of cycling and sports during summer and winter season. [b]Alcohol consumption: low: men: ⩽1 glass/day, women ⩽½ glass/day; moderate: men: 1–2 glasses/day; women ½–1 glass/day; high: men: >2 glasses/day; women >1 glass/day. [c]Sum of concentrations of *n* individual metabolites for each subclass.

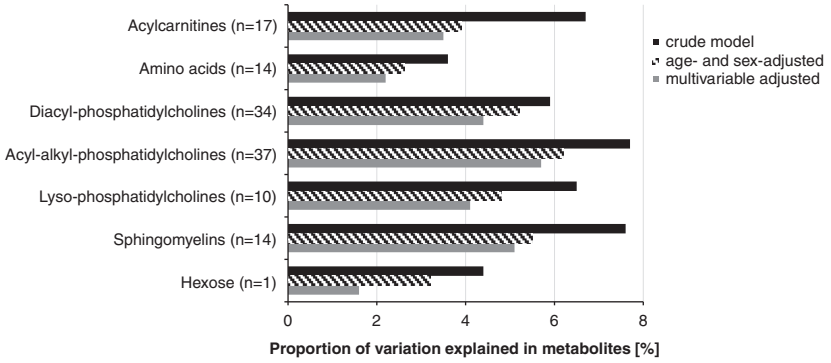

Figure 1. Explained variation of metabolites by maximum number of diet factor scores as obtained from RRR comparing different adjustment models in the EPIC-Potsdam subcohort ($n = 2380$). The multivariable model was adjusted for age (years), sex, alcohol intake from beverages (non-consumers; women: >0–6 g/day, 6–12 g/day, >12 g/day; men: >0–12 g/day, 12–24 g/day, >24 g/d), smoking (never, former, current $\leqslant 20$ cigarettes/day, current >20 cigarettes/day), physical activity (cycling and sports in hour/week), education (low, medium and high), prevalent hypertension, prevalent diabetes, BMI (kg/m^2) and waist circumference (cm). Note: the maximum number of factor scores is given by the number of individual metabolites per metabolite subclass (*n*).

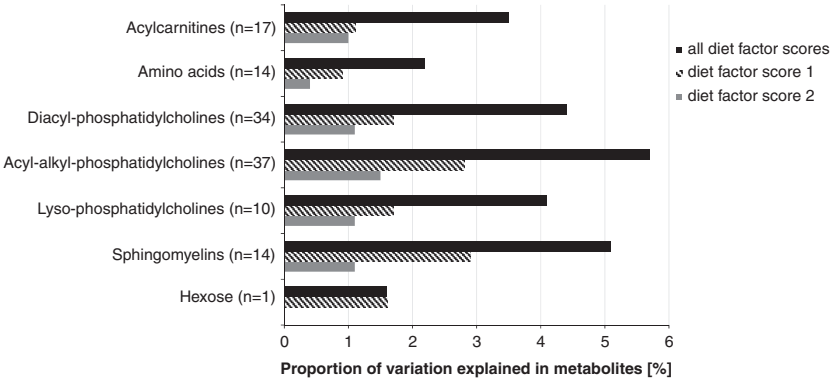

Figure 2. Explained variation of metabolites by maximum number of diet factor scores and first two diet factor scores as obtained from RRR for the multivariable-adjusted model in the EPIC-Potsdam subcohort ($n = 2380$). The multivariable model was adjusted for age (years), sex, alcohol intake from beverages (non-consumers; women: >0–6 g/day, 6–12 g/day, >12 g/day; men: >0–12 g/day, 12–24 g/day, >24 g/day), smoking (never, former, current $\leqslant 20$ cigarettes/day, current >20 cigarettes/day), physical activity (cycling and sports in hour/week), education (low, medium and high), prevalent hypertension, prevalent diabetes, BMI (kg/m^2) and waist circumference (cm). Note: the maximum number of factor scores is given by the number of individual metabolites per metabolite subclass (*n*).

lyso-phosphatidylcholine C17:0. A second pattern with high intake of margarine, non-whole-grain bread, meat and coffee and low intake of butter, pasta/rice and tea was particularly linked to lyso-phosphatidylcholines C20:4 and C18:2 (Table 3). The class of sphingomyelins was related to a dietary pattern with high intake of butter, garlic and coffee and low intake of margarine, fresh fruit and soup (Table 2). This pattern was particularly related to the hxdroxy-sphingomyelins. A pattern with low intake of butter, sweet foods, high-fat dairy, fruits and whole-grain bread was particularly associated with the sphingomyelins without hydroxy group (Table 3). In a sensitivity analysis, the main analysis was stratified by sex. The proportion of variation of habitual diet related to metabolites was higher for men than women across all metabolite classes. Important foods, however, were similar between both sexes (data not shown).

In a hypothesis-driven approach we used two metabolite factors as response variables that have been linked to risk of type 2 diabetes in this population previously.[7] The explained variation of the metabolite factors by dietary patterns was small in the multivariable-adjusted model ($<2\%$). A dietary pattern with high intake of red meat and low intake of whole-grain bread and tea was particularly linked to a metabolite factor associated with higher risk of type 2 diabetes (Table 4). A pattern consisting of high intake of coffee, cake and cookies, canned fruit and fish was linked to a metabolite factor associated with lower risk of type 2 diabetes.

DISCUSSION

In the present study, we observed that overall only a small proportion of the variation of serum metabolites, including acylcarnitines, amino acids, hexose, diacyl-, acyl-alkyl- and lyso-phosphatidylcholines and sphingomyelins, was related to habitual diet in the frame of a population-based study. The proportion of explained variation was of similar magnitude as reported in previous studies that investigated dietary patterns in relation to inflammatory biomarkers and blood lipids.[27,28] In general, this observation may point towards a minor role of usual diet as a determinant of changes in these serum metabolites; however, as food intake was assessed by FFQ, there is a chance of measurement error that could lead to underestimation of the true effects. Diet, in general, may be a direct source of metabolites; however, the interpretation of blood metabolite concentrations remains complex, because other processes have to be considered,

Table 2. Important foods and the individual metabolites that were best explained by the first diet factor score

Metabolite class	*Important foods*[a]		*Individual metabolites*[b] *(% variation explained)*
	Positive loading	*Negative loading*	
Acylcarnitines	Butter (0.68)	Margarine (−0.54), low-fat cheese (−0.20)	C8:1 (−5.0), C9 (4.2), C18:2 (−3.8), C16 (1.7), C18 (1.4), C14:2 (−0.9), C10 (0.3), C10:2 (−0.3), C2 (0.3), C18:1 (0.2), C14:1 (0.1), C3 (−0.1), C0 (−0.1), C5-DC/C6-OH (0), C16:2 (0), C7-DC (0), C5-OH/C3-DCM (0)
Amino acids	Potatoes (0.33), desserts (0.29), cornflakes, crisps (0.27), low-fat dairy products (0.26), raw vegetables (0.25), high-fat dairy products (0.23)	Soup (−0.34), high-energy soft drinks (−0.21), fruit juice (−0.20)	Met (1.6), Ile (1.3), Thr (1.3), Val (1.2), Trp (1.2), His (1.1), Tyr (0.9), Arg (0.8), Gln (0.8), Phe (0.7), Orn (0.7), Pro (0.5), Gly (0.4), Ser (0.3)
Hexose	Red Meat (0.26), non-whole grain bread (0.26), fish (0.25), eggs (0.20)	Whole-grain bread (−0.42), tea (−0.37), butter (−0.20)	Hexose (1.7)
Diacyl-phosphatidylcholines	Fish (0.26), poultry (0.20)	Confectionary (−0.33), cake, cookies (−0.26), desserts (−0.25), margarine (−0.25), tea (−0.24), whole grain bread (−0.23), pasta, rice (−0.21), high-fat cheese (−0.20)	C38:6 (6.1), C36:0 (5.4), C38:0 (5.1), C40:6 (4.0), C36:5 (3.7), C36:6 (3.3), C42:2 (3.1), C42:5 (2.6), C38:5 (2.4), C42:6 (2.0), C34:1 (1.8), C38:1 (1.7), C40:3 (1.6), C32:0 (1.6), C36:4 (1.6), C36:1 (1.3), C42:1 (1.3), C40:2 (1.2), C38:4 (1.1), C42:0 (0.8), C42:4 (0.8), C40:5 (0.8), C32:1 (0.7), C34:4 (0.7), C34:2 (0.6), C36:2 (0.4), C30:0 (0.3), C34:3 (0.2), C36:3 (0.2), C38:3 (0.2), C40:4 (0.2), C32:3 (0.1), C28:1 (0), C32:2 (0)
Acyl-alkyl-phosphatidylcholines	Butter (0.61), red meat (0.22), high-fat dairy products (0.20)	Margarine (−0.44), whole-grain bread (−0.22)	C34:0 (10.7), C36:1 (9.7), C34:1 (8.1), C30:0 (7.8), C40:2 (6.0), C38:4 (5.2), C38:3 (4.7), C38:1 (4.2), C36:2 (4.2), C32:2 (3.3), C36:4 (3.2), PC40:5 (3.2), C32:1 (3.1), C34:2 (3.0), C36:0 (2.8), C40:6 (2.6), C38:6 (2.4), C38:5 (2.2), C36:5 (2.1), C40:3 (2.0), C36:3 (2.0), C38:2 (2.0), C42:2 (1.9), C40:4 (1.9), C30:2 (1.4), C30:1 (1.1), C38:0 (0.6), C34:3 (0.5), C40:1 (0.4), C44:6 (−0.3), C42:5 (0.3), C44:4 (0.2), C42:4 (0.2), C44:5 (0.1), C42:3 (0.1), C44:3 (0), C42:1 (0)
Lyso-phosphatidylcholines	Butter (0.54), high-fat dairy products (0.30), sweet bread spreads (0.22)	Margarine (−0.44), low-fat cheese (−0.23), pasta, rice (−0.23)	C17:0 (6.8), C14:0 (2.0), C18:1 (2.0), C28:1 (2.0), C16:1 (1.4), C20:3 (1.1), C16:0 (0.6), C20:4 (0.5), C18:0 (0.4), C18:2 (0.1)
Sphingomyelins	Butter (0.49), garlic (0.21), coffee (0.20)	Margarine (−0.51), fresh fruit (−0.25), soup (−0.21)	OH-C16:1 (8.8), OH-C14:1 (8.0), OH-C24:1 (4.6), OH-C22:2 (4.6), OH-C22:1 (4.2), C26:0 (2.9), C18:0 (2.6), C18:1 (1.4), C16:0 (1.1), C26:1 (0.6), C16:1 (0.5), C24:0 (0.4), C20:2 (−0.3), C24:1 (0)

[a]Presented are foods with factor loading $\geqslant|0.2|$; factor loadings are presented in parenthesis. Model was multivariable-adjusted for age (years), sex, alcohol intake from beverages (non-consumers; women: >0–6 g/day, 6–12 g/day, >12 g/day; men: >0–12 g/day, 12–24 g/day, >24 g/day), smoking (never, former, current $\leqslant$20 cigarettes/day and current >20 cigarettes/day), physical activity (cycling and sports in h/week), education (low, medium and high), BMI (kg/m^2), waist circumference (cm), prevalent hypertension (*y*/*n*) and prevalent type 2 diabetes (*y*/*n*). [b]Metabolites are ranked from largest to smallest % variation explained by the diet factor score. C*x*:*y*, where *x* represents the number of carbon atoms and *y* the number of double bonds of fatty acids. C0 is free carnitine. DC, dicarboxyl; OH, hydroxy; M, methyl. Negative sign indicates negative weight of response variable.

such as metabolic transformations of nutrients, gut microbiota composition, homeostatic control mechanisms and *de novo* synthesis, to name a few.[8] It is possible, that the metabolites included into this study may better reflect endogenous processes than dietary intake. In addition, recent diet may be a stronger determinant of serum metabolites compared with habitual diet.[29] Of the metabolite classes, choline-containing phospholipids and acylcarnitines were most explained by dietary patterns in our study. These classes include lipid metabolites that are involved in fatty acid transport and oxidation,[1,30] and thus, could be reflective of fat intake and consequent metabolism. Hexose and amino-acid concentrations were least explained by dietary patterns. This may be due to the tight regulation of blood levels of glucose and amino acids by homeostasis and hormonal control mechanisms.[31,32] In the following paragraphs, the individual metabolite classes are discussed in more detail.

Acylcarnitines

Acylcarnitines represent the fatty-acid derivates of carnitine and are substrates and products of mitochondrial beta-oxidation. Free carnitine in blood can directly be derived from dietary

Table 3. Important foods and individual metabolites that were best explained by the second diet factor score

Metabolite class	*Important foods*[a]		*Individual metabolites*[b] *(% variation explained)*
	Positive loading	*Negative loading*	
Acylcarnitines	Cornflakes, crisps (0.20)	Fish (−0.28), other vegetable fat (−0.27), whole grain bread (−0.24), cooked vegetables (−0.23), garlic (−0.23), nuts (−0.23), tea (−0.23), cabbage (−0.21), sweet bread spreads (−0.21), cake, cookies (−0.20), high-fat cheese (−0.20)	C14:1 (2.02), C2 (1.97), C18 (1.84), C14:2 (1.83), C18:1 (1.70), C7-DC (1.30), C16 (0.99), C16:2 (0.95), C6-OH/C5-DC (0.81), C10 (0.78), C18:2 (0.54), C8:1 (0.46), C0 (0.40), C10:2 (0.40), C5-OH/C3-DCM (0.17), C9 (−0.06), C3 (0.05)
Amino acids	Canned fruit (0.40), fried potatoes (0.25), legumes (0.23), cake, cookies (0.20)	Water (−0.32), low-fat cheese (−0.29), fish (−0.24), whole grain bread (−0.20), grain flakes, muesli (−0.20)	Ser (1.34), Tyr (−1.02), Gly (0.67), Thr (0.65), Val (−0.61), His (0.25), Ile (−0.21), Pro (−0.13), Gln (0.02), Met (−0.01), Phe (0), Orn (0), Trp (0), Arg (0)
Diacyl-phosphatidylcholines	Sauce (0.24), butter (0.20)	Fish (−0.58), whole grain bread (−0.30), tea (−0.22), grain flakes, muesli (−0.20)	C36:1 (3.67), C28:1 (3.51), C34:1 (3.08), C30:0 (3.00), C32:1 (2.73), C40:4 (2.18), C42:2 (−1.57), C32:2 (1.44), C40:5 (1.42), C34:4 (1.37), C36:0 (−1.37), C42:0 (−1.32), C38:0 (−1.28), C42:1 (−1.23), C32:0 (1.23), C36:3 (1.04), C34:3 (1.03), C40:3 (−0.99), C38:3 (0.79), C40:2 (−0.67), C38:6 (−0.62), C38:5 (0.58), C36:4 (0.56), C38:1 (−0.50), C38:4 (0.49), C40:6 (−0.41), C32:3 (0.35), C36:2 (0.06), C36:5 (−0.04), C42:4 (0.03), C42:6 (0.03), C34:2 (0.01), C42:5 (0), C36:6 (0)
Acyl-alkyl-phosphatidylcholines	Red meat (0.44), processed meat (0.32), poultry (0.29), margarine (0.21), non-whole-grain bread (0.21)	Sweet bread spreads (−0.24), butter (−0.24), desserts (−0.22), high-fat dairy products (−0.21), tea (−0.20), vegetarian dishes (−0.20)	C36:4 (7.43), C38:5 (7.18), C36:5 (6.57), C38:6 (4.04), C30:0 (−3.89), C36:3 (3.22), C36:1 (−2.96), C34:0 (−2.86), C34:3 (2.78), C34:1 (−1.56), C34:2 (1.49), C38:4 (1.24), C42:1 (1.08), C44:6 (0.81), C40:1 (0.66), C38:3 (−0.64), C44:5 (0.63), C42:5 (0.62), C40:4 (0.61), C36:2 (−0.61), C30:2 (−0.57), C40:2 (−0.56), C30:1 (−0.34), C40:5 (0.31), C44:3 (0.30), C40:3 (−0.23), C36:0 (0.22), C42:4 (0.18), C42:3 (0.17), C38:1 (−0.16), C40:6 (0.07), C38:2 (0.05), C32:1 (0.05), C38:0 (0.02), C42:2 (−0.01), C32:2 (−0.01), C44:4 (0)
Lyso-phosphatidylcholines	Margarine (0.34), non-whole grain bread (0.31), processed meat (0.29), red meat (0.25), coffee (0.25)	Butter (−0.32), pasta, rice (−0.31), tea (−0.21), desserts (−0.21), soup (−0.20)	C20:4 (2.91), C18:2 (2.19), C18:0 (1.36), C20:3 (1.26), C17:0 (−0.97), C28:1 (−0.79), C16:0 (0.65), C18:1 (0.63), C14:0 (−0.56), C16:1 (−0.11)
Sphingomyelins		Butter (−0.45), sweet bread spreads (−0.31), high-fat cheese (−0.27), fresh fruit (−0.26), whole grain bread (−0.25), desserts (−0.25), cake, cookies (−0.24), high-fat dairy products (−0.24)	C24:1 (4.41), OH-C14:1 (−1.95), C24:0 (1.95), C26:1 (1.70), C16:0 (1.45), C18:0 (1.44), C18:1 (1.21), C16:1 (1.11), OH-C16:1 (−0.75), OH-C22:2 (−0.09), OH-C24:1 (−0.03), C26:0 (0.01), C20:2 (0), OH-C22:1 (0)

Note: there is only one response variable for hexose; thus, only one diet factor score was extracted. [a]Presented are foods with factor loading $\geqslant|0.2|$; factor loadings are presented in parenthesis. Model was multivariable-adjusted for age (years), sex, alcohol intake from beverages (non-consumers; women: >0–6 g/day, 6–12 g/day, >12 g/day; men: >0–12 g/day, 12–24 g/day, >24 g/day), smoking (never, former, current ⩽20 cigarettes/day and current >20 cigarettes/day), physical activity (cycling and sports in hour/week), education (low, medium and high), BMI (kg/m^2), waist circumference (cm), prevalent hypertension (*y*/*n*) and prevalent type 2 diabetes (*y*/*n*). [b]Metabolites are ranked from largest to smallest % variation explained by the diet factor score. C*x*:*y*, where *x* represents the number of carbon atoms and *y* the number of double bonds of fatty acids. C0 is free carnitine. DC, dicarboxyl; OH, hydroxy; M, methyl. Negative sign indicates negative weight of response variable.

sources, particularly red meat, as well as secondary from *de novo* synthesis.[33] In the present study, saturated acylcarnitines were linked to butter intake and unsaturated acylcarnitines to margarine intake. This may be explained by the differential fatty acid composition of butter and margarine, as butter usually contains higher amounts of saturated fatty acids.[34] Thus, our observations suggest that serum acylcarnitines may reflect degradation products from fatty acid oxidation which were initially supplied from fatty food intake.

Amino acids

In our study, methionine and branched chain amino acids were explained by a pattern high in dairy products, potatoes, vegetables and cornflakes, and low in beverages. Serine was better explained by a dietary pattern characterized by high intake of canned fruit, fried potatoes and legumes and low intake of water and low-fat cheese. In agreement, important dietary sources for amino acids include meat, dairy products, legumes, starchy foods and cereals, and to lower amounts, fruits and vegetables.[35,36] In our study, in general the essential amino acids were better explained by habitual diet, which is plausible; thereby, methionine was best explained. Methionine conversion generates the universal methyl-group donor S-adenosylmethionine, which is part of many biochemical pathways and also required for epigenetic modifications.[37] Interestingly, methionine also represents an intermediate of carnitine and phosphatidylcholine biosynthesis and is thereby connected to lipid metabolism.[38] Methionine may be particularly derived from cereals and it has been shown that methionine may be protective in atherosclerosis and cognitive impairment.[38] The branched chain amino acids isoleucine and valine may be directly derived from dietary sources. In addition, they have been linked to higher risk of type 2 diabetes in the Framingham Offspring study and Malmö Diet and Cancer study[6] and recently also in EPIC-Potsdam.[7] Serine, however, is not only derived from diet but also from *de novo* synthesis.

Hexose

Hexose represents the sum of all six-carbon monosaccharides; thereby, the major component is glucose. Recent work from our

Table 4. Dietary patterns associated with metabolite factors which have been linked to risk of type 2 diabetes

Important foods[a]		*Metabolite factor*[b]
Positive loading	*Negative loading*	
Red meat (0.30) Poultry (0.27) Butter (0.22) Other vegetable fat (0.20)	Whole grain bread (-0.38), tea (-0.32), margarine (-0.28), soup (-0.24), pasta, rice (-0.24)	Metabolite factor 2 associated with higher risk of type 2 diabetes
Coffee (0.41) Cake, cookies (0.36) Canned fruit (0.30) Fish (0.29) Butter (0.29)	Low-fat dairy products (-0.25)	Metabolite factor 1 associated with lower risk of type 2 diabetes

[a]Presented are foods with factor loading $\geqslant|0.2|$; factor loadings are presented in parenthesis. Model was multivariable-adjusted for age (years), sex, alcohol intake from beverages (non-consumers; women: >0–6 g/day, 6–12 g/day, >12 g/day; men: >0–12 g/day, 12–24 g/day, >24 g/day), smoking (never, former, current $\leqslant 20$ cigarettes/day and current >20 cigarettes/day), physical activity (cycling and sports in hour/week), education (low, medium and high), BMI (kg/m^2), waist circumference (cm), prevalent hypertension (*y*/*n*) and prevalent type 2 diabetes (*y*/*n*). [b]Two metabolite factors that have previously been derived by principal component analysis and that were linked to risk of incident type 2 diabetes in this population[7] were included as response variables, and consequently two dietary patterns were derived. The metabolite factors were calculated as linear combinations of the weighted and standardized metabolite concentrations according to the following formulas: factor $1 = 0.82 \times$ PC aa C42:0 $+ 0.79 \times$ PC aa C42:1 $+ 0.80 \times$ PC ae C32:1 $+ 0.78 \times$ PC ae C32:2 $+ 0.70 \times$ PC ae C34:2 $+ 0.72 \times$ PC ae C34:3 $+ 0.71 \times$ PC ae C36:2 $+ 0.71 \times$ PC ae C36:3 $+ 0.85 \times$ PC ae C40:5 $+ 0.76 \times$ PC ae C40:6 $+ 0.82 \times$ PC ae C42:3 $+ 0.85 \times$ PC ae C42:4 $+ 0.87 \times$ PC ae C42:5 $+ 0.76 \times$ PC ae C44:4 $+ 0.78 \times$ PC ae C44:5 $+ 0.83 \times$ PC ae C44:6 $+ 0.54 \times$ SM C16:1 $+ 0.57 \times$ SM OH-C22:2 $+ 0.41 \times$ lysoPC a C17:0. Factor $2 = 0.55 \times$ propionyl-carnitine $+ 0.66 \times$ phenylalanine $+ 0.61 \times$ tryptophan $+ 0.66 \times$ tyrosine $+ 0.68 \times$ valine $+ 0.66 \times$ isoleucine $+ 0.59 \times$ PC aa C32:1 $+ 0.70 \times$ PC aa C36:1 $+ 0.65 \times$ PC aa C36:3 $+ 0.76 \times$ PC aa C38:3 $+ 0.72 \times$ PC aa C40:4 $+ 0.71 \times$ PC aa C40:5 $+ 0.44 \times$ hexose.

group suggested that hexose concentrations were strongly linked to risk of type 2 diabetes.[7] Hexose in our study was linked to a pattern with high intake of red meat and non-whole-grain bread and low intake of whole-grain bread and tea. It has previously been reported that a 'traditional' dietary pattern containing red meat was associated with higher levels of blood glucose in a Dutch population.[39] Accordingly, some intervention studies reported higher blood glucose levels with increased red meat intake, whereas others reported no association.[40,41] Furthermore, consumption of red meat and high-glycemic index foods has previously been found to be associated with higher risk of chronic diseases such as type 2 diabetes.[42] In addition, a recent meta-analysis reported that high whole-grain intake was associated with lower risk of type 2 diabetes and CVD in prospective cohort studies, and lower plasma glucose concentrations in intervention studies.[43] Tea intake has also been inversely associated with risk of chronic diseases.[44] With hexose as an established marker of hyperglycemia a plausible dietary pattern including foods that have previously been linked to risk of chronic disease could be identified. This underlines the potential of the statistical method RRR to identify biologically relevant dietary patterns, and encourages its use also for less-established biomarkers and metabolites of chronic disease risk.

Choline-containing phospholipids

Phosphatidylcholines, lyso-phosphatidylcholines and sphingomyelins are main components of cellular membranes and part of blood lipoproteins. They can be distinguished according to the type of bonding (diacyl-phosphatidylcholines contain two ester-bonds whereas acyl-alkyl-phosphatidylcholines contain an ester and an ether bond), number of fatty acid bonds (lyso-phosphatidylcholines are derived from phosphatidylcholines by elimination of one fatty acid) and backbone (sphingomyelins contain a ceramid core instead of glycerol, which is the backbone in phosphatidylcholines). In our study, diacyl-phosphatidylcholines were positively linked to fish and poultry intake and inversely associated with the intake of sweet foods, whole-grain bread and tea. Poultry and fish are important sources of dietary choline, which is required for hepatic phosphatidylcholine biosynthesis.[45,46] It was previously shown that mice on a choline-deficient diet had lower phosphatidylcholine concentrations, and accumulated hepatic fat.[47] Particularly, diacyl-phosphatidylcholines C36:1, C32:1 and C40:4, which were linked to a dietary pattern characterized by low intake of fish, whole-grain bread and tea and high intake of sauce, have been linked to higher risk of type 2 diabetes in this population previously.[7] Acyl-alkyl-phosphatidylcholines, sphingomyelins and lyso-phosphatidylcholines were linked to high butter and high-fat dairy and low margarine intake. A previous meta-analysis including intervention studies on butter versus margarine intake reported that these foods differentially affected blood lipoproteins, which could also affect the phospholipids, and concluded that high-quality margarine should be preferred over butter.[48] In our study population, however, acyl-alkyl-phosphatidylcholines, sphingomyelins and lyso-phosphatidylcholines, which were linked to a pattern with high butter intake, were linked to lower risk of type 2 diabetes.[7] Thus, butter intake may not necessarily be linked to higher risk of chronic diseases. An individual metabolite that was largely explained by a dietary pattern with high intake of butter and high-fat dairy and low margarine intake in our study was lysophosphatidylcholine C17:0, which contains heptadecanoic acid. This saturated fatty acid with odd number of carbon atoms cannot be synthesized by the human body, but only by microbiota in the rumen of ruminants. Thus, it is mainly derived from milk fat, and was previously suggested as a biomarker for dairy fat intake.[49] This may be one reason why it may better be explained by habitual diet than other metabolites in our study.

Metabolite factors associated with risk of type 2 diabetes

We identified type 2 diabetes-relevant dietary patterns with two metabolite factors as response variables. Important foods included high intake of red meat, low intake of whole-grain bread, tea, coffee, cake and cookies, canned fruit and fish. Most of these foods have previously been associated with higher risk of chronic diseases in this study population and other prospective cohorts.[24,42–44]

The strengths of our study include that we conducted a metabolomics analysis in the frame of a large population-based study. We considered a large variety of foods and metabolites, which underscores the exploratory nature of our study. Further, we applied a statistical method from dietary pattern analysis to a metabolomics study, which to our knowledge has not been done, yet. Our study was, however, limited. We used a cross-sectional design, and therefore, cannot show temporality and causality. However, little is known about the association between habitual

diet and variation of serum metabolites on population scale; thus, this cross-sectional analysis may give some first ideas. In addition, it may be a particular challenge to conduct a prospective study on this topic, as prospective cohort studies are often restricted to a single baseline blood sample collection. Habitual diet was self-reported and assessed by FFQ, which is a feasible instrument in epidemiologic studies. Although the EPIC-Potsdam FFQ included a large variety of foods, has previously been validated and was shown to generate reproducible results,[16–19] it may be susceptible to measurement error. In general, imprecision of the dietary assessment could lead to underestimation of the impact of habitual diet on serum concentrations of metabolites.[50] In addition, a systematic measurement error could lead to over- or underestimation of the true effect. To overcome this limitation, another approach would be to assess recent diet by use of 24-hour dietary recalls, and to study the association between recent diet and serum metabolites in future studies. Measurement of serum metabolites may depend on the analytical technique used, and may be prone to high within-person variation. However, as shown in our previous study,[22] most serum metabolites were reliable over time; metabolites that showed high technical variation were further excluded. Last, the study is limited to the metabolites included into the kit. It is possible that other metabolite classes may be more strongly linked to habitual diet, for example, secondary plant metabolites with fruit and vegetable intake.

In conclusion, in the frame of a population-based study we found that habitual diet was weakly linked to serum metabolites studied here. Dietary patterns that contributed most to variation of serum metabolites were characterized by the following foods: butter, margarine, whole-grain bread, tea, coffee, red meat, fish and dairy products. Of them, particularly, dietary patterns with high intake of red meat and low intake of whole-grain bread, tea and coffee were linked to disease-relevant metabolites, and thus, could be potential targets for chronic disease prevention. Future studies are needed to verify these findings. They should include additional metabolites and more precise dietary assessment methods, for example, 24- hour dietary recalls.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We specially thank Ellen Kohlsdorf and Wolfgang Bernigau for data management. We also thank Werner Römisch-Margl, Julia Scarpa, and Arsin Sabunchi for metabolomics measurements performed at the Helmholtz Centrum München, Genome Analysis Center, Metabolomics Core Facility. We are thankful to all the EPIC-Potsdam study participants for their devoted participation in the study. This study was supported by the Federal Ministry of Science, Germany (grant no. 01 EA 9401), the European Union (grant no. SOC 95 201408 05F02) and the German Cancer Aid (grant no. 70-2201-Bo2), and by a grant from the Federal Ministry of Education and Research (BMBF) to the German Center for Diabetes Research (DZD eV) (Förderkennzeichen: 01G10922).

REFERENCES

1 Adams SH, Hoppel CL, Lok KH, Zhao L, Wong SW, Minkler PE *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.
2 Suhre K, Meisinger C, Doring A, Altmaier E, Belcredi P, Gieger C *et al.* Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. *PLoS One* 2010; **5**: e13953.
3 Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y *et al.* Novel biomarkers for pre-diabetes identified by metabolomics. *Mol Syst Biol* 2012; **8**: 615.
4 Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF *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.
5 Fiehn O, Garvey WT, Newman JW, Lok KH, Hoppel CL, Adams SH. Plasma metabolomic profiles reflective of glucose homeostasis in non-diabetic and type 2 diabetic obese African-American women. *PLoS One* 2010; **5**: e15234.
6 Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E *et al.* Metabolite profiles and the risk of developing diabetes. *Nat Med* 2011; **17**: 448–453.
7 Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG *et al.* Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes* 2013; **62**: 639–648.
8 Gibney MJ, Walsh M, Brennan L, Roche HM, German B, van Ommen B. Metabolomics in human nutrition: opportunities and challenges. *Am J Clin Nutr* 2005; **82**: 497–503.
9 Primrose S, Draper J, Elsom R, Kirkpatrick V, Mathers JC, Seal C *et al.* Metabolomics and human nutrition. *Br J Nutr*. 2011; **105**: 1277–1283.
10 Hyotylainen T, Bondia-Pons I, Oresic M. Lipidomics in nutrition and food research. *Mol Nutr Food Res* 2013; **57**: 1306–1318.
11 Altmaier E, Kastenmuller G, Romisch-Margl W, Thorand B, Weinberger KM, Illig T *et al.* Questionnaire-based self-reported nutrition habits associate with serum metabolism as revealed by quantitative targeted metabolomics. *Eur J Epidemiol* 2011; **26**: 145–156.
12 Hoffmann K, Schulze MB, Schienkiewitz A, Nothlings U, Boeing H. Application of a new statistical method to derive dietary patterns in nutritional epidemiology. *Am J Epidemiol* 2004; **159**: 935–944.
13 Boeing H, Korfmann A, Bergmann MM. Recruitment procedures of EPIC-Germany. European Investigation into Cancer and Nutrition. *Ann Nutr Metab* 1999; **43**: 205–215.
14 Boeing H, Wahrendorf 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.
15 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* 2008; **31**: 1138–1143.
16 Kroke A, Klipstein-Grobusch K, Voss S, Moseneder J, Thielecke F, Noack R *et al.* Validation of a self-administered food-frequency questionnaire administered in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study: comparison of energy, protein, and macronutrient intakes estimated with the doubly labeled water, urinary nitrogen, and repeated 24-h dietary recall methods. *Am J Clin Nutr* 1999; **70**: 439–447.
17 Boeing H, Bohlscheid-Thomas S, Voss S, Schneeweiss S, Wahrendorf J. The relative validity of vitamin intakes derived from a food frequency questionnaire compared to 24-hour recalls and biological measurements: results from the EPIC pilot study in Germany. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol* 1997; **26**(Suppl 1): S82–S90.
18 Bohlscheid-Thomas S, Hoting I, Boeing H, Wahrendorf J. Reproducibility and relative validity of energy and macronutrient intake of a food frequency questionnaire developed for the German part of the EPIC project. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol* 1997; **26**(Suppl 1): S71–S81.
19 Bohlscheid-Thomas S, Hoting I, Boeing H, Wahrendorf J. Reproducibility and relative validity of food group intake in a food frequency questionnaire developed for the German part of the EPIC project. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol* 1997; **26**(Suppl 1): S59–S70.
20 Romisch-Margl W, Prehn C, Bogumil R, Rohring C, Suhre K, Adamski J. Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. *Metabolomics* 2012; **8**: 133–142.
21 US Department of Health and Human Services, FDA, Center for Drug Evaluation and Research (CDER), Center for Veterinary Medicine (CVM). Guidance for industry. *Bioanalytical Method Validation* 2001.
22 Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J *et al.* Reliability of serum metabolite concentrations over a 4-month period using a targeted metabolomic approach. *PLoS One* 2011; **6**: e21103.
23 Schulze MB, Hoffmann K, Kroke A, Boeing H. Dietary patterns and their association with food and nutrient intake in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study. *Br J Nutr* 2001; **85**: 363–373.
24 von Ruesten A, Feller S, Bergmann MM, Boeing H. Diet and risk of chronic diseases: results from the first 8 years of follow-up in the EPIC-Potsdam study. *Eur J Clin Nutr* 2013; **67**: 412–419.
25 Schulze MB, Hoffmann K, Manson JE, Willett WC, Meigs JB, Weikert C *et al.* Dietary pattern, inflammation, and incidence of type 2 diabetes in women. *Am J Clin Nutr* 2005; **82**: 675–684 (quiz 714–5).
26 Fung TT, Schulze MB, Hu FB, Hankinson SE, Holmes MD. A dietary pattern derived to correlate with estrogens and risk of postmenopausal breast cancer. *Breast Cancer Res Treat* 2012; **132**: 1157–1162.

27 Meyer J, Doring A, Herder C, Roden M, Koenig W, Thorand B. Dietary patterns, subclinical inflammation, incident coronary heart disease and mortality in middle-aged men from the MONICA/KORA Augsburg cohort study. *Eur J Clin Nutr* 2011; **65**: 800–807.
28 McNaughton SA, Mishra GD, Brunner EJ. Food patterns associated with blood lipids are predictive of coronary heart disease: the Whitehall II study. *Br J Nutr* 2009; **102**: 619–624.
29 O'Sullivan A, Gibney MJ, Brennan L. Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. *Am J Clin Nutr* 2011; **93**: 314–321.
30 Quehenberger O, Dennis EA. The human plasma lipidome. *N Engl J Med* 2011; **365**: 1812–1823.
31 McBride KL, Belmont JW, O'Brien WE, Amin TJ, Carter S, Lee BH. Heritability of plasma amino acid levels in different nutritional states. *Mol Genet Metab* 2007; **90**: 217–220.
32 Hers HG. Mechanisms of blood glucose homeostasis. *J Inherit Metab Dis* 1990; **13**: 395–410.
33 Pekala J, Patkowska-Sokola B, Bodkowski R, Jamroz D, Nowakowski P, Lochynski S *et al.* L-carnitine—metabolic functions and meaning in humans life. *Curr Drug Metab* 2011; **12**: 667–678.
34 US Department of Agriculture, Agricultural Research Service. 2012. USDA National Nutrient Database for Standard Reference, Release 25. Nutrient Data Laboratory Homepagehttp://www.ars.usda.gov/ba/bhnrc/ndl (accessed: 14th January 2013).
35 FAO. Nutritional studies: amino-acid content of foods and biological data on proteins. *FAO Nutr Stud* 1970; **24**: 1–285.
36 Schonfeldt HC, Gibson Hall N. Dietary protein quality and malnutrition in Africa. *Br J Nutr* 2012; **108**(Suppl 2): S69–S76.
37 Zeisel SH. Epigenetic mechanisms for nutrition determinants of later health outcomes. *Am J Clin Nutr* 2009; **89**: 1488S–1493SS.
38 Fardet A. New hypotheses for the health-protective mechanisms of whole-grain cereals: what is beyond fibre? *Nutr Res Rev* 2010; **23**: 65–134.
39 van Dam RM, Grievink L, Ocke MC, Feskens EJ. Patterns of food consumption and risk factors for cardiovascular disease in the general Dutch population. *Am J Clin Nutr* 2003; **77**: 1156–1163.
40 Smith DR, Wood R, Tseng S, Smith SB. Increased beef consumption increases apolipoprotein A-I but not serum cholesterol of mildly hypercholesterolemic men with different levels of habitual beef intake. *Exp Biol Med (Maywood)* 2002; **227**: 266–275.
41 Murphy KJ, Thomson RL, Coates AM, Buckley JD, Howe PR. Effects of eating fresh lean pork on cardiometabolic health parameters. *Nutrients* 2012; **4**: 711–723.
42 Pan A, Sun Q, Bernstein AM, Schulze MB, Manson JE, Willett WC *et al.* Red meat consumption and risk of type 2 diabetes: 3 cohorts of US adults and an updated meta-analysis. *Am J Clin Nutr* 2011; **94**: 1088–1096.
43 Ye EQ, Chacko SA, Chou EL, Kugizaki M, Liu S. Greater whole-grain intake is associated with lower risk of type 2 diabetes, cardiovascular disease, and weight gain. *J Nutr* 2012; **142**: 1304–1313.
44 Jing Y, Han G, Hu Y, Bi Y, Li L, Zhu D. Tea consumption and risk of type 2 diabetes: a meta-analysis of cohort studies. *J Gen Intern Med* 2009; **24**: 557–562.
45 Cole LK, Vance JE, Vance DE. Phosphatidylcholine biosynthesis and lipoprotein metabolism. *Biochim Biophys Acta* 2011; **1821**: 754–761.
46 Li Z, Vance DE. Phosphatidylcholine and choline homeostasis. *J Lipid Res* 2008; **49**: 1187–1194.
47 Raubenheimer PJ, Nyirenda MJ, Walker BR. A choline-deficient diet exacerbates fatty liver but attenuates insulin resistance and glucose intolerance in mice fed a high-fat diet. *Diabetes* 2006; **55**: 2015–2020.
48 Zock PL, Katan MB. Butter, margarine and serum lipoproteins. *Atherosclerosis* 1997; **131**: 7–16.
49 Wolk A, Vessby B, Ljung H, Barrefors P. Evaluation of a biological marker of dairy fat intake. *Am J Clin Nutr* 1998; **68**: 291–295.
50 Schatzkin A, Kipnis V. Could exposure assessment problems give us wrong answers to nutrition and cancer questions? *J Natl Cancer Inst* 2004; **96**: 1564–1565.

Supplementary Information accompanies this paper on European Journal of Clinical Nutrition website (http://www.nature.com/ejcn)