



HAL
open science

Questionnaire-based self-reported nutrition habits associate with serum metabolism as revealed by quantitative targeted metabolomics

Elisabeth Altmaier, Gabi Kastenmüller, Werner Römisch-Margl, Barbara Thorand, Klaus M. Weinberger, Thomas Illig, Jerzy Adamski, Angela Döring,
Karsten Suhre

► To cite this version:

Elisabeth Altmaier, Gabi Kastenmüller, Werner Römisch-Margl, Barbara Thorand, Klaus M. Weinberger, et al.. Questionnaire-based self-reported nutrition habits associate with serum metabolism as revealed by quantitative targeted metabolomics. *European Journal of Epidemiology*, 2010, 26 (2), pp.145-156. 10.1007/s10654-010-9524-7. hal-00646466

HAL Id: hal-00646466

<https://hal.science/hal-00646466>

Submitted on 30 Nov 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Questionnaire-based self-reported nutrition habits associate with serum metabolism as revealed by quantitative targeted metabolomics

Elisabeth Altmaier^{1,2,*}, Gabi Kastenmüller¹, Werner Römisch-Margl¹, Barbara Thorand³, Klaus M. Weinberger⁴, Thomas Illig³, Jerzy Adamski^{5,6}, Angela Döring³, Karsten Suhre^{1,2}

¹ *Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany*

² *Faculty of Biology, Ludwig-Maximilians-Universität, 82152 Planegg-Martinsried, Germany*

³ *Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany*

⁴ *Biocrates Life Sciences AG, 6020 Innsbruck, Austria*

⁵ *Institute of Experimental Genetics, Genome Analysis Center, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany*

⁶ *Institute of Experimental Genetics, Life and Food Science Center Weihenstephan, Technische Universität München, 85354 Freising-Weihenstephan, Germany*

** To whom correspondence should be addressed. Helmholtz Zentrum München, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany, e-mail: elisabeth.altmaier@helmholtz-muenchen.de, Phone: ++49-89-3187-3640, Fax: ++49-89-3187-3585*

ABSTRACT

Nutrition plays an important role in human metabolism and health. However, it is unclear in how far self-reported nutrition intake reflects de facto differences in body metabolite composition. To investigate this question on an epidemiological scale we conducted a metabolomics study analyzing the association of self-reported nutrition habits with 363 metabolites quantified in blood serum of 284 male participants of the KORA population study, aged between 55 and 79 years. Using data from an 18-item food frequency questionnaire, the consumption of 18 different food groups as well as four derived nutrition indices summarizing these food groups by their nutrient content were analyzed for association with the measured metabolites. The self-reported nutrition intake index “polyunsaturated fatty acids” associates with a decrease in saturation of the fatty acid chains of glycerophosphatidylcholines analyzed in serum samples.

Using a principal component analysis dietary patterns highly associating with serum metabolite concentrations could be identified. The first principal component, which was interpreted as a healthy nutrition lifestyle, associates with a decrease in the degree of saturation of the fatty acid moieties of different glycerophosphatidylcholines.

In summary, this analysis shows that on a population level metabolomics provides the possibility to link self-reported nutrition habits to changes in human metabolic profiles and that the associating metabolites reflect the self-reported nutritional intake. Moreover, we could show that the strength of association increases when composed nutrition indices are used. Metabolomics may, thus, facilitate evaluating questionnaires and improving future questionnaire-based epidemiological studies on human health.

Keywords: dietary pattern, metabolomics, mass spectrometry, nutrition habits, food questionnaires, epidemiology

Abbreviations:

KORA Cooperative Health Research in the Region of Augsburg;

MONICA Monitoring of Trends and Determinants in Cardiovascular Disease;

DP dietary pattern

INTRODUCTION

Potential causes of complex diseases, such as type 2 diabetes and cardiovascular disease, can be analyzed from many different perspectives (*Figure 1*). For instance, a number of genome-wide association studies (GWAS) identified associations between common genetic polymorphisms (SNPs) and diabetes or coronary artery disease [1-3]. In addition to such heritable genetic predispositions, environmental factors such as the human lifestyle and nutrition habits play an important role in the development of the respective clinical outcome. Examples would be the effect of meat and fish consumption on the risk of colorectal cancer, metabolic differences in the intake of green and black tea or the metabolic effects of dietary isoflavones in humans [4-11]. However, directly associating genes or lifestyle habits with clinical endpoints provides only limited information about the underlying disease-causing mechanisms. Moreover, the p-values of such associations are most often relatively low. Using quantitative targeted metabolomics as an access to intermediate phenotypes (as described in detail in [12]) provides a much closer insight into the mechanisms that cause the clinical endpoint. Using high resolution electrospray ionization tandem mass spectrometry (ESI-MS/MS) it is, today, possible to measure hundreds of metabolites at the same time in a single sample [13]. Gieger et al. [12] showed that associating genetic variants with changes in the homeostasis of metabolic compounds can yield larger effect sizes as well as access to the underlying molecular disease-causing mechanisms. Likewise, lifestyle parameters such as smoking habits and coffee consumption can be clearly discerned in metabolic profiles. *Wang-Sattler et al.* [14] suggest that smoking is associated with plasmalogen-deficiency disorders and *Altmaier et al.* [15] could identify an increase in the concentration of two classes of sphingomyelins and a decrease in the acylcarnitine levels with an increased coffee consumption.

Here, we focus on the effect of nutrition habits on the human metabolism and in particular on how self-reported nutrition intake associates with metabolic profiles of the study participants' blood samples. Epidemiological studies provide growing evidence that diet is associated with the risk of many chronic diseases. Thus, an accurate assessment of nutrition habits is of growing importance. Though traditional measurement methods are improving and new techniques are developed [16-20], the complexity of nutrition intake naturally leads to errors in its assessment. Comprehensive information about nutrition habits would include a multitude of different factors, such as portion size of foods, frequency of consumption, food composition, daily variations in food intake and many more. However, documenting all these parameters in epidemiological studies is not always feasible. Thus, errors in the assessment of the consumed food are to be expected [21, 22] and an independent validation against quantitatively measurable parameters is needed.

Here, we ask the question, in how far self-reported nutrition habits based on questionnaires can be associated with changes in the metabolic homeostasis. For clinical short-term studies, where factors like food intake are exactly defined and controlled, finding such associations would probably be not astonishing. Here, however, we focus on population studies, where a general nutrition habit is self-reported. Analyzing population studies for associations between the self-reported nutrition habits and the metabolic profile might then evaluate the quality and potential of the used questionnaires.

We present results from a study where we measured the metabolic profiles of 284 male participants from the KORA (Cooperative Health Research in the Region of Augsburg) study population. For each participant, information on many different factors concerning disease state, medication items or life style is available. In particular, most participants responded to a set of 18 questions of a qualitative food frequency questionnaire (see *Table 4*). Each of the queried food groups such as meat, fish, vegetables, cheese or chocolate was reported on a scale from 1 ("never") to 7 ("several times per day"). In a first step we analyzed these KORA

food groups for associations with the measured metabolic profiles. In a second step we defined nutrition indices by pooling the food groups by their nutrient content, such as “dietary fibers”, “saturated fatty acids”, “monounsaturated fatty acids”, and “polyunsaturated fatty acids” and tested them for associations with the metabolic data. In a third and final step we examined the effect of the overall diet on the human metabolism. For identifying associations of the nutrition lifestyle with the clinical outcome, dietary pattern analysis is a common approach [23-25]. Following this idea, we identified dietary patterns and analyzed them for associations with the measured metabolic profiles. This approach may provide a broader picture of the impact of specific nutrition habits on health than the analysis of single food groups could do alone.

MATERIAL & METHODS

Experimental Setup

The research platform KORA (Cooperative Health Research in the Region of Augsburg) conducts population-based surveys and subsequent follow-up studies in the fields of health care research, epidemiology, health economics and genetics. A multitude of different parameters is provided, including life style factors (questionnaires related to physical activity, nutrition, alcohol consumption, smoking, etc.), medical history, and socio-demographic variables. The dataset analyzed here was taken from the F3 study, which was conducted in 2004-2005 as a follow-up of the third MONICA (Monitoring of Trends and Determinants in Cardiovascular Disease) survey (S3; 1994/1995). For this follow-up study F3, 3006 of the 4856 participants of the S3 were reexamined for a second time, thus, 10 years apart from the baseline examination. From this group, 284 male participants, aged between 55 and 79 years at the examination for F3, were chosen randomly for our analysis and were again recruited one to two years later for additional blood sampling. In total, 239 of these study participants

provided details concerning their intake of the following foods: “Meat (except sausage)”, “sausages and ham”, “poultry”, “fish”, “potatoes”, “pasta”, “rice”, “cooked vegetables”, “fresh fruit”, “chocolate, chocolates”, cakes, pastries, biscuits”, “salted snacks”, “whole-grain bread, black bread, crispbread”, “flaked oats, muesli, cornflakes”, “curds, yoghurt and sour milk”, “cheese”, “eggs”, and “milk”. The foods were queried by a score from 1 to 7: “never”, “once a month or less”, “several times a month”, “about once a week”, “several times a week”, “daily or almost daily”, “several times per day” (*see Table 1*). Food indices were computed as the sum of the responses that make up the given indices (*see Table 4*). Thus, the indices “dietary fibers” and “polyunsaturated fatty acids” can each range from 3 to 21, “monounsaturated fatty acids” from 8 to 56, and “saturated fatty acids” ranges from 9 to 63. This food frequency questionnaire was validated against a 7-day dietary record (n=899; applied in a previous KORA/MONICA study) with the result that it can be used for analysis on group level [26].

Blood samples

For collection of blood samples for metabolic analysis F3 study participants were invited again in 2006. To avoid variation due to circadian rhythm, blood was drawn in the morning between 8 and 10 am after a period of overnight fasting. Material was immediately horizontally shaken (10 min), followed by 40 min resting at 4°C to obtain complete coagulation. The material was then centrifuged (2000g; 4°C). Serum was aliquoted and stored for 2-4 hours at 4°C, after which it was deep frozen to -80°C until mass spectrometry analysis.

Metabolite profiling

Targeted metabolite profiling by electrospray ionization (ESI) tandem mass spectrometry (MS/MS) was performed at *Biocrates life sciences AG*, Austria. The technique is described in

detail in [27, 28]. A comprehensive overview of the field and the related technologies is given in the review paper by Wenk [29]. Briefly, the assay preparation was done by an automated robotics system (Hamilton Robotics GmbH) on a special double-filter 96 well plate containing isotope labeled internal standards. Assays used 10 μ l plasma and include phenylisothiocyanate (PITC)-derivatisation of amino acids, extraction with organic solvent and several liquid handling steps. Flow injection analysis coupled with multiple reaction monitoring scans (FIA MS/MS) on an API 4000 QTrap instrument (Applied Biosystems) was used for quantification of amino acids, acylcarnitines, sphingomyelins, phosphatidylcholines, and hexoses. LC-MS/MS methods using multiple reaction monitoring, neutral loss and precursor ion scans were applied for biogenic amines, eicosanoids and hydroxylated fatty acid derivatives as well as for intermediates of the energy metabolism. The quantification of the metabolites is achieved by reference to appropriate internal standards. The method is proven to be in conformance with the “Guidance for Industry - Bioanalytical Method Validation” published by the FDA (Food and Drug Administration), which implies proof of reproducibility within a given error range. This measurement platform has been successfully used in the past in different academic and industrial applications [12-14].

Metabolite spectrum

For all analyzed metabolites the concentrations are reported in μ M. In total, 363 different metabolites were measured in plasma: 18 amino acids, nine reducing mono-, di- and oligosaccharides (abbreviated as H n for n -hexose, dH for desoxyhexose, UA for uronic acid, HNAc for N-acetylglucosamine), 21 acylcarnitines (C x : y , where x denotes the number of carbons in the side chain and y the number of double bonds), seven hydroxylacylcarnitines (C(OH) x : y) and dicarboxylacylcarnitines (C x : y -DC), free carnitine (C0), seven biogenic amines, seven prostaglandins and 293 lipids. These lipids are subdivided into 14 different ceramides (Cer) and glucosylceramides (GlcCer), 71 different sphingomyelins (SM x : y) and

sphingomyelin-derivatives, such as N-hydroxydicarboacyloylsphingosyl-phosphocholine (SM(OH,COOH)_{x:y}) and N-hydroxyacyloylsphingosyl-phosphocholine (SM (OH)_{x:y}), five glycerophosphatidic acids (PA), 124 glycerophosphatidylcholines (PC), 42 glycerophosphatidylethanolamines (PE), four phosphatidylglycerols (PG), 30 glycerophosphatidylinositols (PI) and glycerophosphatidylinositol-bisphosphate (PIP₂), and three glycerophosphatidylserines (PS). Glycerophospholipids are further differentiated with respect to the presence of ester (*a*) and ether (*e*) bonds in the glycerol moiety, where two letters (*aa*, *ae*, or *ee*) denote that the first as well as the second position of the glycerol unit are bound to a fatty acid residue, while a single letter (*a* or *e*) indicates a bond with only one fatty acid residue; the latter molecular species are usually called lyso-phospholipids. E.g. PC *ae* 36:1 denotes a plasmalogen phosphatidylcholine with 36 carbons in the two fatty acid side chains and a single double bond in one of them. In some cases, the mapping of metabolite names to individual masses can be ambiguous. For example, stereo-chemical differences are not always discernable, neither are isobaric fragments.

Statistical analysis

The statistical analysis system R (<http://www.r-project.org/>) and SPSS for Windows (Version 16.0, Chicago: SPSS Inc.) were used for this statistical analysis.

As a basis for our statistical analysis, we used in particular ratios between metabolite concentrations. As shown previously, the use of such ratios leads to a strong reduction in the overall variance and therefore improves the strength of association [13]. A high increase in the strength of association may also indicate that the two metabolites are linked by a metabolic pathway that is influenced by the respective (nutrition) factor.

Thus, the ratios of the concentrations of all possible pairs (363^2) of metabolites are computed. For the identification of metabolites influenced by the nutrition habits, for each of the 18 food groups, the four nutrition indices and the seven dietary patterns, a linear regression test was

applied to model each metabolite ratio. To control for the effect of testing multiple hypotheses and, thus, taking possible correlations among metabolites into account, the positive false discovery rate (q-value) was computed for each ratio and each dietary variable. The positive false discovery rate is defined by Storey et al. [30] as $pFDR = E[V/R \mid R > 0]$, where “positive” means that at least one rejection occurred. The pFDR can be used to define the q-value, which is a natural pFDR analogue to the p-value. Thus, the q-value is a measure for the fraction of false positives appearing even if the test itself was significant. In contrast to the p-value the q-value is a multiple hypothesis testing quantity [30].

The dietary patterns were identified using principal component analysis on the food groups using SPSS. Only principal components with an eigenvalue larger than one were considered for further analysis. For each of these principal components a new variable with the transformed values (method: regression) was calculated. To each of these new variables a linear regression test was applied to model each metabolite ratio analogical to the food variables above. Thereby, metabolite ratios can be identified that are influenced by the respective dietary pattern.

RESULTS

Food groups associate with saturation and chain length of phospholipids

As a first step, in order to gain insight into nutrition associated metabolic changes, we tested the association of self-reported intake of single food groups with the metabolite profiles in human serum using linear regression. The results are summarized in *Table 2*. Using age and BMI as cofactors for the linear regression showed no substantial difference to the analysis without any cofactors. All analyzed food groups mainly affected the degree of saturation and/or the chain length of the fatty acid residues of the glycerophosphatidylcholines.

An association with a decrease in saturation of the fatty acid residues could be observed for the consumption of “sausages, ham” and “fish” (e.g. the ratio between PC aa C40:4 and PC aa C40:6 with a q-value of 8.67×10^{-10}). In contrast, an increase in saturation was associated with “flaked oats, muesli, cornflakes” and “cheese”. The results for “flaked oats, muesli, cornflakes” additionally revealed a decrease in the chain lengths of the fatty acid residues. For “vegetables, cooked” the ratio between hexanoylcarnitine (C6) and kynurenin concentrations was found as association with the lowest q-value (q-value: 5.89×10^{-3}). “Fresh fruit” associated most strongly with a ratio composed of an acylcarnitine and a glucosylceramide, namely C10/GlcCer C22:2 (q-value: 2.27×10^{-3}) and “chocolate, chocolates” with ratios containing each a sphingomyelin and a glycerophosphatidylcholine, e.g. SM C22:0/PC aa C38:3 (q-value: 1.35×10^{-4}). “Curds, yoghurt, sour milk” achieved the lowest q-value for association with SM (OH) C28:1/SM (OH) C20:1 (q-value: 2.93×10^{-2}), a ratio composed of two N-hydroxyacyloylsphingosyl-phosphocholines. For all these associations positive false discovery rates (q-value) below 0.05 were thus obtained. For the following food groups no association with a positive false discovery rate (q-value) smaller than 0.05 could be identified: “meat (except sausage)”, “poultry”, “potatoes”, “pasta”, “rice”, “cakes, pastries, biscuits”, “salted snacks”, “whole-grain bread, black bread, crispbread”, “eggs” and “milk”. In summary, the analysis of the food groups shows that associations of several of these food groups with different metabolic profiles can be observed, albeit some of them with relatively low q-values. Most of these differences are related to the saturation and chain length of phospholipids.

The nature of the nutrient intake indices matches the character of the associating metabolic profiles

As a next step, we aimed to identify those metabolites that are influenced by the food indices “dietary fibers”, “saturated fatty acids”, “monounsaturated fatty acids”, and “polyunsaturated fatty acids”.

The index “dietary fibers” was defined by the food groups “vegetables, cooked”, “whole-grain bread, black bread, crispbread” and “flaked oats, muesli, cornflakes” (*Table 4*). For this index “dietary fibers” metabolite ratios that suggest a shift towards more saturation and smaller chain length of the fatty acid residues showed the strongest association (*Table 3*), e.g. the ratio PC ae C36:1/ PC aa C40:4 with a positive false discovery rate (q-value) of 1.20×10^{-5} and a positive correlation coefficient of 0.4. In order to investigate this observed shift towards more saturation further, sums of the serum concentrations of glycerophosphatidylcholines were used for analysis. The association with the lowest q-value (2.45×10^{-7}) was the ratio of the sum of all acyl-alkyl-PC with one or two double bonds and the sum of all diacyl-PC with 4 or 5 double bonds (*Figure 2*). This supports the idea that “dietary fibers” are associated with an increase in phospholipid saturation. To answer the question whether the binding type of the fatty acids in the glycerophosphatidylcholine – acyl or alkyl – may influence the observed association, we focused on those ratios that contain exclusively diacyl-PCs (*Table 3b*) or exclusively acyl-alkyl-PCs (*Table 3c*). For both separately analyzed classes also an increase of the saturation of the fatty acid residues could be observed. The sum of all diacyl-PC showed only a slight negative association with dietary fibers (p-value: 1.53×10^{-3} , q-value: 1.79×10^{-2}), while for the sum of all acyl-alkyl-PC no association could be detected. Thus, the intake of dietary fibers mainly associates with the saturation of the fatty acid residues but might also have an effect on the type of PC that is formed.

The index “dietary fibers” is the sum of the food groups “cooked vegetables”, “whole-grain bread, black bread, crispbread” and “flaked oats, muesli, cornflakes” (*Table 4*). The analysis

of these factors showed that only self-reported “flaked oats, muesli, cornflakes” (q-value: 2.24×10^{-4}) consumption was also associated with a higher saturation and a smaller fatty acid chain length. The analysis of the index “dietary fibers” lead to an increased strength of association and, thus, shows that additional information can be obtained when using nutrition indices as compared to individual food groups.

The index “polyunsaturated fatty acids” is composed of the three food groups: “meat (except sausage)”, “poultry” and “fish”. This index was associated with a decrease in saturation of the fatty acid residues of the glycerophosphatidylcholines, where the signal of the strongest associating ratio is found for PC aa C40:5/PC aa C40:6 with a small positive false discovery rate (q-value: 5.94×10^{-5}) and a negative correlation coefficient of -0.38 (*Figure 3*). Docosapentaenoyl-CoA (C22:5) and Docosahexaenoyl-CoA (C22:6) are closely connected in the biosynthesis of unsaturated fatty acids [31]. Synthesis of these metabolites to a glycerol 3-phosphate, and further addition of a stearyl-moiety (C18:0), followed by a dephosphorylation step and the addition of a phosphocholin moiety in the Kennedy pathway, leads to the formation of the glycerol-phosphatidylcholins PC aa C40:5 and PC aa C40:6, respectively. Thus, a decrease of the ratio PC aa C40:5/PC aa C40:6 indicates an increase in the polyunsaturated fatty acid C22:6. The index “polyunsaturated fatty acids” is defined by the food groups “meat (except sausage)”, “poultry” and “fish”. Among these groups the strongest association could be found for fish intake with the ratio PC aa C40:4/PC aa C40:6 (q-value: 8.67×10^{-10}).

The indices “saturated fatty acids” and “monounsaturated fatty acids” are both composed of the eight food groups “meat (except sausage)”, “sausages and ham”, “poultry”, “chocolate, chocolates”, “cakes, pastries, biscuits”, “salted snacks”, “cheese” and “eggs”. This nearly identical composition of these both indices is typical for southern German (Bavarian) lifestyle [32]. The index “saturated fatty acids” in addition contains the food group “curds, yoghurt,

sour milk”. Both indices associated with metabolite ratios composed of the different phospholipids.

Three of the index defining food groups, namely “sausages and ham”, “chocolate, chocolates” and “cheese”, achieved results with a q-value lower than 5%. The resulting ratios mainly describe the state of saturation of the fatty acid residues of the glycerophosphatidylcholines.

The index “saturated fatty acids” includes, besides the eight food groups, additionally the factor “curds, yoghurt and sour milk”. This difference to “monounsaturated fatty acids” became apparent in additionally resulting ratios containing the sphingomyelin SM (OH) C20:1, which was, among the food groups, only included in the results of “curds, yoghurt and sour milk”.

The indices “saturated fatty acids” and “monounsaturated fatty acids” present a complex pattern of differences in the phospholipid metabolism, which is probably due to the large number of food compounds defining these indices suggesting that these indices may need refinement in the future. In summary, especially the analysis of the index “polyunsaturated fatty acids” provides an impressive example that the self-reported nutrient intake can in fact be observed in the metabolite profile of human serum.

Dietary lifestyle can be detected in serum metabolite profile

To identify the major dietary patterns in the KORA F3 dataset, we analyzed the food groups by principal component analysis (*Table 4*). Seven of the resulting principal components achieved an eigenvalue larger than one and were further analyzed. Each of the seven principal components can be interpreted as a dietary pattern (DP). The association of these seven DPs with the metabolite dataset was tested. Three of the seven DPs yielded results with a false discovery rate (q-value) lower than 0.05. The lowest q-value was achieved by the first principal component. This first component was composed of a high intake of poultry, fish, rice, cooked vegetables, fresh fruit, chocolate, flaked oats, curds, cheese and milk, and a low

consumption of meat and sausages. The metabolite ratios that most strongly associate with this DP consisted each of a diacyl-PC with four or five double bonds in the numerator and an acyl-alkyl-PC with one or two double bonds in the denominator and showed a positive correlation. This fact is particularly visible when examining the ratio of the sums of these both groups of glycerophosphatidylcholines for analysis (q-value: 1.07×10^{-8}) (Figure 4, Table 5). The fourth and the sixth DP were both associated with different phospholipids. The fourth DP describes a lifestyle that incorporates a high consumption of meat, poultry, fish, salted snacks and eggs and a low intake of potatoes, pasta, cakes, whole-grain bread and flaked oats (lowest q-value for association with metabolite ratios: 3.42×10^{-3}). It can be interpreted as “traditional Bavarian” lifestyle, since meat, sausages, poultry and fish played an important role in the nutrition (note that the probands live in Bavaria). The sixth DP indicates a high intake of whole-grain bread, eggs and milk, and a low consumption of fish, fresh fruit and cakes (lowest q-value: 1.35×10^{-5}). These three principal components (component one, four and six) explain a variance of 15.6, 7.9% and 6.0% respectively. Because of its high explained variance the first principal component was of main interest. The first DP associates with a decrease in saturation of the fatty acid side chains of glycerophosphatidylcholines. Thus, the principal component that explains most of the variance in the self-reported dietary intake is also the variable that displays the strongest association to serum measured profiles.

DISCUSSION

In this study we asked the question if self-reported nutrition intake reflects de facto changes in body metabolite composition. We analyzed the effect of different nutrition habits on the metabolism of 239 male participants of the KORA project, aged between 55 and 79 years. In total, 363 metabolites were measured for each sample by high-throughput electrospray ionization tandem mass spectrometry.

Our analysis shows that self-reported information about general nutrition habits collected by food frequency questionnaires can be associated with different metabolic profiles. This could be done for a population study, which means that food intake was not predetermined as it is for clinical short-term studies. Apart from that the associations could be found though the time points of the interview and the blood collection differed by a year. This, in reverse, indicates that nutrition habits are stable over longer time periods. We could find associations between self-reported nutrition intake and serum based metabolite concentrations on three levels of analysis: single food groups, nutrition indices and dietary patterns. Using age and BMI as cofactors for the linear regression showed no substantial difference to the analysis without any cofactors. Life-style factors as well as possible gene-nutrition interactions may influence the relationship between nutrition intake and metabolic profile, but were not used as cofactors for reasons of possible over-fitting of the statistical model.

Analyzing the 18 single food groups we showed that even on the lowest level, namely single food groups, self-reported information about these single food consumptions can be associated with different metabolic profiles. Most of these differences relate to the saturation and chain length of different classes of phospholipids.

In a next step we combined single food groups into indices according to their nutrient content and showed that nutrition habits can be associated with metabolic changes and thereby provide additional information. For the index “dietary fibers”, we observed an association with a shift towards more saturation and smaller chain length of the fatty acid residues of phosphatidylcholins. To the best of our knowledge, such an effect of dietary fiber consumption on the saturation of fatty acid residues in humans was not reported in literature so far. Thus, the observed association was unexpected and confounding by other aspects of diet can not totally be excluded. An explanation for the association might be that increased dietary fiber intake functions as an indicator for increased carbohydrate intake, which was observed to be directly correlated to adipose tissue monounsaturated fatty acids [33]. This

association between dietary fiber intake and saturation was stronger than any of the associations between the food compounds that build this index and the metabolic data. This shows that additional information can be extracted from the data by using nutrition indices compared to food groups. For instance, the index “polyunsaturated fatty acids” was associated with a decrease in saturation of the fatty acid side chains of the glycerophosphatidylcholines. This provides a convincing example that the self-reported increased intake of polyunsaturated fatty acids can be observed in the serum as an increase in glycerophosphatidylcholines containing polyunsaturated fatty acids. Applying a principal component analysis to all food groups to identify dietary patterns the effect of the overall diet on the human metabolism was shown. The dietary pattern with the largest explained variance can be compared to a similar pattern found by *Liu et al.* [25] in a multi-ethnic US population, who denote this pattern as “health conscious”. This lifestyle may be interpreted as “healthy”, basically since meat and sausages are replaced by poultry and fish. It was strongly associated with a – as healthy considered – decrease in saturation of the fatty acid side chains of glycerophosphatidylcholines. Thus, the self-reported healthy lifestyle including a high content of consumed fish and poultry can effectively be detected in the human blood and this even months or years after the report.

The associations we reported in this paper were found in a metabolite set that contained no free fatty acids, but glycerophospholipids. It is true that using free fatty acids would provide additional information. Here, however, we use a high-throughput technology that allows the measurement of hundreds of metabolites in a fully automated manner for many samples at a time which is more appropriate for future large population studies. Moreover, the different phospholipid species are not only meant as substitutes for free fatty acid profiles, but carry much independent information e.g. on lipid trafficking, membrane composition and others [29]. Also discussable is that as study participants only older men were chosen who may not

be as good informed about their nutrition intake as their wives. This selection is due to the fact that our study originated from an explorative metabolomics study with multiple objectives. However, the fact that we still obtain a significant signal under more unfavourable conditions supports the validity of our results.

For our analysis we used information about food frequencies and not about portion sizes of food. Of course, using portion sizes would be a more precise way to assess nutrition intake. However, in the setting of an epidemiological examination, the questionnaires are limited by the very nature of this kind of large scale population studies. The fact, that we obtain significant results despite this limitation shows again the power of this approach.

In summary, we have shown that metabolomics provides the possibility to evaluate and improve questionnaires and to detect which single foods should be chosen to compose a meaningful nutrition index. This, again, may improve future questionnaire-based studies on human health. Beyond this, metabolomics might be a future tool to identify biomarkers reflecting the effect of nutrition intake on the human metabolism. Such dietary biomarkers could then be an objective and independent assessment for the “true” food intake [34-36]. In addition, other dietary assessment methods (than questionnaires) [37-39] could be validated using metabolic biomarkers.

ACKNOWLEDGEMENTS

The KORA research platform and the MONICA Augsburg studies were initiated and financed by the Helmholtz Zentrum München, - National Research Center for Environmental Health, which is funded by the German Federal Ministry of Education, Science, Research and Technology and by the State of Bavaria. The KORA study group consists of H.-E. Wichmann (speaker), R. Holle, J. John, T. Illig, C. Meisinger, A. Peters, and their coworkers, who are responsible for the design and conduct of the KORA studies. We gratefully acknowledge the contribution of all members of field staffs who were involved in planning and conducting the MONICA/KORA Augsburg studies. Finally, we express our appreciation to all study participants for donating their blood and time.

This work was partially funded by the Federal Ministry of Education and Research within the SysMBo project (project number: 0315494A), by the research consortium GANI_MED (project number: 03IS2061A) and by the Deutsches Zentrum für Diabetesforschung e.V..

REFERENCES

1. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet.* 2008 May;40(5):638-45.
2. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet.* 2007 Jul;39(7):857-64.
3. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, et al. Genomewide association analysis of coronary artery disease. *N Engl J Med.* 2007 Aug 2;357(5):443-53.
4. Norat T, Bingham S, Ferrari P, Slimani N, Jenab M, Mazuir M, et al. Meat, fish, and colorectal cancer risk: the European Prospective Investigation into cancer and nutrition. *J Natl Cancer Inst.* 2005 Jun 15;97(12):906-16.
5. Van Dorsten FA, Daykin CA, Mulder TP, Van Duynhoven JP. Metabonomics approach to determine metabolic differences between green tea and black tea consumption. *J Agric Food Chem.* 2006 Sep 6;54(18):6929-38.
6. Walsh MC, Brennan L, Pujos-Guillot E, Sebedio JL, Scalbert A, Fagan A, et al. Influence of acute phytochemical intake on human urinary metabolomic profiles. *Am J Clin Nutr.* 2007 Dec;86(6):1687-93.
7. Solanky KS, Bailey NJ, Beckwith-Hall BM, Bingham S, Davis A, Holmes E, et al. Biofluid 1H NMR-based metabonomic techniques in nutrition research - metabolic effects of dietary isoflavones in humans. *J Nutr Biochem.* 2005 Apr;16(4):236-44.
8. Kimura Y, Kono S, Toyomura K, Nagano J, Mizoue T, Moore MA, et al. Meat, fish and fat intake in relation to subsite-specific risk of colorectal cancer: The Fukuoka Colorectal Cancer Study. *Cancer Sci.* 2007 Apr;98(4):590-7.
9. Chao A, Thun MJ, Connell CJ, McCullough ML, Jacobs EJ, Flanders WD, et al. Meat consumption and risk of colorectal cancer. *JAMA.* 2005 Jan 12;293(2):172-82.
10. Cross AJ, Leitzmann MF, Gail MH, Hollenbeck AR, Schatzkin A, Sinha R. A prospective study of red and processed meat intake in relation to cancer risk. *PLoS Med.* 2007 Dec;4(12):e325.
11. Stella C, Beckwith-Hall B, Cloarec O, Holmes E, Lindon JC, Powell J, et al. Susceptibility of human metabolic phenotypes to dietary modulation. *J Proteome Res.* 2006 Oct;5(10):2780-8.
12. Gieger C, Geistlinger L, Altmaier E, Hrabce de Angelis M, Kronenberg F, Meitinger T, et al. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* 2008 Nov;4(11):e1000282.
13. Altmaier E, Ramsay SL, Graber A, Mewes HW, Weinberger KM, Suhre K. Bioinformatics analysis of targeted metabolomics--uncovering old and new tales of diabetic mice under medication. *Endocrinology.* 2008 Jul;149(7):3478-89.
14. Wang-Sattler R, Yu Y, Mittelstrass K, Lattka E, Altmaier E, Gieger C, et al. Metabolic profiling reveals distinct variations linked to nicotine consumption in humans--first results from the KORA study. *PLoS ONE.* 2008;3(12):e3863.
15. Altmaier E, Kastenmuller G, Romisch-Margl W, Thorand B, Weinberger KM, Adamski J, et al. Variation in the human lipidome associated with coffee consumption as revealed by quantitative targeted metabolomics. *Mol Nutr Food Res.* 2009 Oct 6.
16. Dowell SA, Welch JL. Use of electronic self-monitoring for food and fluid intake: A pilot study. *Nephrol Nurs J.* 2006 May-Jun;33(3):271-7.
17. Kikunaga S, Tin T, Ishibashi G, Wang DH, Kira S. The application of a handheld personal digital assistant with camera and mobile phone card (Wellnavi) to the general population in a dietary survey. *J Nutr Sci Vitaminol (Tokyo).* 2007 Apr;53(2):109-16.

18. Subar AF, Thompson FE, Potischman N, Forsyth BH, Buday R, Richards D, et al. Formative research of a quick list for an automated self-administered 24-hour dietary recall. *J Am Diet Assoc.* 2007 Jun;107(6):1002-7.
19. Slimani N, Valsta L. Perspectives of using the EPIC-SOFT programme in the context of pan-European nutritional monitoring surveys: methodological and practical implications. *Eur J Clin Nutr.* 2002 May;56 Suppl 2:S63-74.
20. Wang DH, Kogashiwa M, Kira S. Development of a new instrument for evaluating individuals' dietary intakes. *J Am Diet Assoc.* 2006 Oct;106(10):1588-93.
21. Michels KB. The role of nutrition in cancer development and prevention. *Int J Cancer.* 2005 Mar 20;114(2):163-5.
22. Jenab M, Slimani N, Bictash M, Ferrari P, Bingham SA. Biomarkers in nutritional epidemiology: applications, needs and new horizons. *Hum Genet.* 2009 Jun;125(5-6):507-25.
23. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol.* 2002 Feb;13(1):3-9.
24. Fung TT, Schulze M, Manson JE, Willett WC, Hu FB. Dietary patterns, meat intake, and the risk of type 2 diabetes in women. *Arch Intern Med.* 2004 Nov 8;164(20):2235-40.
25. Liu L, Nettleton JA, Bertoni AG, Bluemke DA, Lima JA, Szklo M. Dietary pattern, the metabolic syndrome, and left ventricular mass and systolic function: the Multi-Ethnic Study of Atherosclerosis. *Am J Clin Nutr.* 2009 Jun 10.
26. Winkler G, Doring A. Validation of a short qualitative food frequency list used in several German large scale surveys. *Z Ernahrungswiss.* 1998 Sep;37(3):234-41.
27. Weinberger KM. [Metabolomics in diagnosing metabolic diseases]. *Ther Umsch.* 2008 Sep;65(9):487-91.
28. Weinberger KM, Graber A. Using Comprehensive Metabolomics to Identify Novel Biomarkers. *Screening Trends in Drug Discovery.* 2005;6:42-5.
29. Wenk MR. The emerging field of lipidomics. *Nat Rev Drug Discov.* 2005 Jul;4(7):594-610.
30. Storey JD. The positive false discovery rate: A Bayesian interpretation and the q-value. *The Annals of Statistics* 2003;31(6):2013-35.
31. Sprecher H, Luthria DL, Mohammed BS, Baykousheva SP. Reevaluation of the pathways for the biosynthesis of polyunsaturated fatty acids. *J Lipid Res.* 1995 Dec;36(12):2471-7.
32. Winkler G, Doring A, Keil U. Trends in dietary sources of nutrients among middle-aged men in southern Germany. Results of the MONICA Project Augsburg: dietary surveys 1984/1985 and 1994/1995. *MONItoring trends and determinants in CARDiovascular disease. Appetite.* 2000 Feb;34(1):37-45.
33. Rubba P, Fidanza F, Gautiero G, Leccia G, Cozzolino G, Mancini M. Influence of dietary intake of energy and carbohydrate on the proportion of saturated and monounsaturated fatty acids in adipose tissue of middle aged men. *Int J Vitam Nutr Res.* 1990;60(4):383-91.
34. Day N, McKeown N, Wong M, Welch A, Bingham S. Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. *Int J Epidemiol.* 2001 Apr;30(2):309-17.
35. Kaaks R, Ferrari P, Ciampi A, Plummer M, Riboli E. Uses and limitations of statistical accounting for random error correlations, in the validation of dietary questionnaire assessments. *Public Health Nutr.* 2002 Dec;5(6A):969-76.
36. Sugar EA, Wang CY, Prentice RL. Logistic regression with exposure biomarkers and flexible measurement error. *Biometrics.* 2007 Mar;63(1):143-51.
37. Bingham SA. Biomarkers in nutritional epidemiology. *Public Health Nutr.* 2002 Dec;5(6A):821-7.
38. Potischman N, Freudenheim JL. Biomarkers of nutritional exposure and nutritional status: an overview. *J Nutr.* 2003 Mar;133 Suppl 3:873S-4S.

39. Tasevska N, Runswick SA, McTaggart A, Bingham SA. Urinary sucrose and fructose as biomarkers for sugar consumption. *Cancer Epidemiol Biomarkers Prev.* 2005 May;14(5):1287-94.

FIGURES AND TABLES

Figure 1: Overview over different perspectives for analyzing complex diseases:

A number of genome-wide association (GWA) studies could identify associations between genetic polymorphisms and diseases. One step closer to the possible causes are the so called intermediate phenotypes (IP) like metabolic profiles. In addition to the genetic background, environmental factors such as the human lifestyle play an important role in the development of the respective clinical outcome. The lifestyle is usually reported by questionnaires, where the information about the nutrition is split into information about food groups. To identify associations between the ingested nutrients and the metabolic profile, the food groups are pooled by their nutrient content to indices.

Figure 2: Dietary fibers associate with an increase in saturation of the fatty acid residues of glycerophosphatidylcholines. Here, we show the boxplots of the plasma concentrations [μM] of the sum of all acyl-alkyl-PC with one or two double bonds on their fatty acid residues divided by the concentrations of the sum of all diacyl-PC with four or five double bonds as a function of dietary fiber intake. The linear regression q-value for the association is 2.45×10^{-7} . Boxes extend from 1st quartile (Q_1) to 3rd quartile (Q_3); median is indicated as a horizontal line; whiskers are drawn to the observation that is closest to, but not more than a distance of $1.5(Q_3 - Q_1)$ from the end of the box. Observations that are more distant than this are shown individually on the plot. The number of individuals in each group is given in the boxes.

Figure 3: Polyunsaturated fatty acid intake associates with a decrease in the concentration of the ratio PC aa C40:5/PC aa C40:6 (legend see Fig. 2; q-value: 5.94×10^{-5}).

Figure 4: First principal component generated by principal component analysis on all food groups associates with a decrease in saturation of the fatty acid residues of glycerophosphatidylcholines. The plasma concentrations [μM] of the sum of all acyl-alkyl-PC with one or two double bonds on their fatty acid residues (PC ae Cx:1,2) divided by the concentrations of the sum of all diacyl-PC with four or five double bonds (PC aa Cx:4,5) is plotted as a function of this first principal component. The linear regression q-value for the association is 1.07×10^{-8} . The linear regression line is shown.

Table 1: Characteristics of the population related to nutrition habits and age. The foods were queried in a food frequency questionnaire by a score from 1 to 7: “never”, “once a month or less”, “several times a month”, “about once a week”, “several times a week”, “daily or almost daily”, “several times per day”.

	Minimum	Maximum	Mean	Standard deviation	Variance
Meat (except sausage)	2	6	3.09	.818	.670
Sausages, ham	1	7	2.86	.932	.868
Poultry	2	7	4.75	1.165	1.357
Fish	1	7	4.69	1.113	1.239
Potatoes	1	7	3.17	.906	.821
Pasta	2	7	3.43	.830	.689
Rice	2	7	4.69	1.219	1.487
Vegetables, cooked	1	7	3.38	.951	.905
Fresh fruit	1	7	2.59	1.178	1.388
Chocolate, chocolates	1	7	4.67	1.738	3.022
Cakes, pastries, biscuits	1	7	4.10	1.467	2.152
Salted snacks	2	7	5.86	1.236	1.527
Whole-grain bread, black bread, crispbread	1	7	3.33	1.747	3.052
Flaked oats, muesli, cornflakes	1	7	5.76	1.732	3.000
Curds, yoghurt, sour milk	1	7	4.07	1.852	3.430
Cheese	2	7	3.09	1.089	1.186
Eggs	2	7	4.67	1.166	1.359
Milk	1	7	4.63	1.978	3.912
Age	55	79	65.91	6.892	47.506

Table 2: Selected results of the linear regression test. For each analyzed factor the associating ratio of metabolites is listed together with the p-value and the positive false discovery rate (q-value). The arrows indicate the direction of the correlation coefficient for each result.

	Metabolites	p-value	q-value		
Food Groups	Meat (except sausage)	–	>0.05		
	Sausages. ham	↓ PC ae C32:0*/ PC aa C40:4 ↑ PC aa C38:4	5.01x10 ⁻⁸ 8.27x10 ⁻⁵	2.20x10 ⁻³ 4.10x10 ⁻³	
	Poultry	–		>0.05	
	Fish	↓ PC aa C40:4/PC aa C40:6	9.91x10 ⁻¹⁵	8.67x10 ⁻¹⁰	
	Potatoes	–		>0.05	
	Pasta	–		>0.05	
	Rice	–		>0.05	
	Vegetables. cooked	↓ C6/Kynurenin	7.58x10 ⁻⁸	5.89x10 ⁻³	
	Fresh fruit	↓ C10/GlcCer C22:2	5.89x10 ⁻⁸	2.27x10 ⁻³	
	Chocolate. chocolates	↑ SM C22:0/PC aa C38:3 ↓ PC aa C38:3/SM (OH.COOH) C18:2	1.43x10 ⁻⁹ 7.53x10 ⁻⁹	1.35x10 ⁻⁴ 3.13x10 ⁻⁴	
	Cakes. pastries. biscuits	–		>0.05	
	Salted snacks	–		>0.05	
	Whole-grain bread. black bread. crispbread	–		>0.05	
	Flaked oats. muesli. cornflakes	↑ PC ae C36:1*/PC aa C38:4	2.91x10 ⁻⁹	2.24x10 ⁻⁴	
	Curds. yoghurt. sour milk	↓ SM (OH) C28:1/SM (OH) C20:1	2.22x10 ⁻⁷	2.93x10 ⁻²	
	Cheese	↓ PC ae C34:6*/PC ae C34:1*	2.06x10 ⁻⁸	7.07x10 ⁻⁴	
	Eggs	–		>0.05	
	Milk	–		>0.05	
	Indices	Dietary fiber	↑ PC ae C36:1*/ PC aa C40:4	9.12x10 ⁻¹¹	1.20x10 ⁻⁵
		Polyunsaturated fatty acids	↓ PC aa C40:5/PC aa C40:6	4.51x10 ⁻¹⁰	5.94x10 ⁻⁵
Monounsaturated fatty acids		↑ SM (OH.COOH) C18:2/ PC aa C36:4	1.09x10 ⁻⁸	1.43x10 ⁻³	
Saturated fatty acids		↑ SM (OH.COOH) C18:2/ PC aa C36:4 ↑ SM (OH) C20:1/PC aa C40:5	2.10x10 ⁻⁸ 3.96x10 ⁻⁸	1.66x10 ⁻³ 1.66x10 ⁻³	
Principal components	Component 1	↑ PC aa C36.4*/PC ae C36:2	6.94x10 ⁻¹³	3.11x10 ⁻⁸	
	Component 4	↓ PC aa C38:0*/Lysine	3.45x10 ⁻⁸	3.31x10 ⁻³	
	Component 6	↓ PC ae (COOH) C30:3/PC aa C40:6	9.92x10 ⁻¹¹	1.31x10 ⁻⁵	

* alternative assignments of the metabolites are possible

Table 3: Results of the linear regression test with the strongest association to dietary fibers (A); limited to those ratios containing only diacyl-PC (B) and those containing only acyl-alkyl-PC (C).

A

Numerator	Denominator	Mean	Standard Deviation	p-value	q-value	Pearson R
PC ae C36:1 *	PC aa C40:4	1.915	0.680	9.1×10^{-11}	1.2×10^{-5}	0.403
PC ae C36:1 *	PC aa C38:4	0.085	0.029	8.6×10^{-10}	2.8×10^{-5}	0.383
PC ae C36:1 *	PC aa C36:4	0.043	0.013	2.9×10^{-9}	6.4×10^{-5}	0.372
PC ae C36:1 *	PC aa C40:5	0.644	0.203	8.8×10^{-9}	1.4×10^{-4}	0.361
PC ae C36:2	PC aa C40:4	3.016	1.247	5.5×10^{-10}	2.5×10^{-5}	0.388
PC ae C36:2	PC aa C38:4	0.133	0.052	5.7×10^{-10}	2.5×10^{-5}	0.387
PC ae C36:2	PC aa C36:4	0.068	0.024	2.7×10^{-9}	6.4×10^{-5}	0.373
PC ae C36:2	PC aa C40:5	1.012	0.365	2.5×10^{-8}	2.6×10^{-4}	0.351
PC ae C34:1 *	PC aa C40:4	2.495	0.886	8.2×10^{-9}	1.4×10^{-4}	0.362
PC ae C32:1 *	PC aa C40:4	1.866	0.665	1.6×10^{-8}	2.1×10^{-4}	0.355
PC ae C32:1 *	PC aa C38:4	0.082	0.028	9.6×10^{-8}	6.3×10^{-4}	0.337
PC ae C38:1 *	PC aa C40:4	1.360	0.529	8.2×10^{-8}	6.0×10^{-4}	0.338
PC ae C38:1 *	PC aa C38:4	0.060	0.021	8.6×10^{-8}	6.0×10^{-4}	0.338

B

Numerator	Denominator	Mean	Standard Deviation	p-value	q-value	Pearson R
PC aa C34:2	PC aa C40:4	94.600	33.152	3.0×10^{-7}	1.0×10^{-3}	0.324
PC aa C34:2	PC aa C38:4	4.186	1.369	3.2×10^{-7}	1.0×10^{-3}	0.323
PC aa C34:2	PC aa C36:4	2.131	0.591	5.5×10^{-7}	1.3×10^{-3}	0.317
PC aa C36:3	PC aa C40:4	28.351	7.324	3.3×10^{-7}	1.1×10^{-3}	0.323
PC aa C36:3	PC aa C38:4	1.261	0.309	6.4×10^{-6}	4.2×10^{-3}	0.287
PC aa C32:2	PC aa C40:4	1.526	0.499	1.8×10^{-6}	2.2×10^{-3}	0.303
PC aa C32:2	PC aa C38:4	0.068	0.022	2.3×10^{-5}	8.1×10^{-3}	0.270
PC aa C36:2	PC aa C40:4	47.536	15.306	1.9×10^{-6}	2.2×10^{-3}	0.303
PC aa C36:2	PC aa C38:4	2.103	0.600	2.6×10^{-6}	2.6×10^{-3}	0.299

C

Numerator	Denominator	Mean	Standard Deviation	p-value	q-value	Pearson R
PC ae C36:2	PC ae C38:3	2.431	0.496	1.2×10^{-6}	1.8×10^{-3}	0.308
PC ae C36:2	PC ae C38:5	0.811	0.260	5.1×10^{-6}	3.7×10^{-3}	0.290
PC ae C36:2	PC ae C36:5 *	1.422	0.513	6.2×10^{-6}	4.0×10^{-3}	0.288
PC ae C36:2	PC ae C36:4	0.760	0.244	8.8×10^{-6}	4.9×10^{-3}	0.283
PC ae C34:2	PC ae C36:5 *	1.291	0.442	8.4×10^{-6}	4.8×10^{-3}	0.284
PC ae C34:2	PC ae C36:4	0.688	0.198	9.1×10^{-6}	5.0×10^{-3}	0.283
PC ae C34:2	PC ae C38:5	0.739	0.230	1.7×10^{-5}	6.8×10^{-3}	0.275
PC ae C34:3	PC ae C36:5 *	0.619	0.190	8.4×10^{-6}	4.8×10^{-3}	0.284

* alternative assignments of the metabolites are possible

Table 4: Composition of the principal components on the food groups and composition of the food indices. Positive values indicate a positive contribution to the respective compound (analogous for negative values). The principal components with an eigenvalue larger than one are listed together with their explained variance. The ‘x’ indicates that the index is deduced from the respective food.

Food groups	Principal components (explained variance in %)							Indices			
	1	2	3	4	5	6	7	Dietary fiber	Polyunsaturated fatty acids	Monounsaturated fatty acids	Saturated fatty acids
	15.693	9.291	8.936	7.967	6.729	6.006	5.565				
Meat (except sausage)	-.264	.370	.465	.300	.239	.070	.262		x	x	x
Sausages. ham	-.290	.048	.647	.261	.134	-.072	.191			x	x
Poultry	.325	.151	-.134	.644	-.261	.014	.093		x	x	x
Fish	.473	.067	-.075	.450	.193	-.348	-.174		x		
Potatoes	.250	.563	.372	-.273	.054	.111	-.039				
Pasta	.253	.547	.307	-.240	-.378	.135	-.175				
Rice	.472	.214	-.298	.150	-.369	.125	-.150				
Vegetables. cooked	.492	.412	.024	-.010	.091	-.172	-.031	x			
Fresh fruit	.579	-.005	.027	.019	.293	-.260	.066				
Chocolate. chocolates	.324	-.436	.523	.010	-.256	.150	-.166			x	x
Cakes. pastries. biscuits	.215	-.332	.457	-.315	-.197	-.244	.051			x	x
Salted snacks	.066	-.502	.268	.310	-.101	-.071	-.335			x	x
Whole-grain bread. black bread. crispbread	.296	-.091	-.123	-.325	.492	.345	-.301	x			
Fleaked oats. muesli. cornflakes	.464	-.204	.016	-.223	.319	-.148	.240	x			
Curds. yoghurt. sour milk	.695	-.059	.006	.004	.064	-.105	.279				x
Cheese	.559	-.075	.160	-.045	-.079	.112	-.144			x	x
Eggs	.191	-.084	.129	.383	.345	.639	-.081			x	x
Milk	.353	-.234	-.124	-.064	-.256	.382	.634				

Table 5: Results of the linear regression test with the strongest association to the first principal component. In cases where alternative assignments of the metabolites are possible, these are indicated by a ‘*’.

Numerator	Denominator	Mean	Standard Deviation	p-value	q-value	Pearson R
PC aa C36:4	Sum(PC ae Cx:1.2)	3.043	0.905	1.52×10^{-13}	1.07×10^{-8}	0.445
Sum(PC aa Cx:4.5)	Sum(PC ae Cx:1.2)	6.289	1.934	2.44×10^{-13}	1.07×10^{-8}	0.442
Sum(PC ae Cx:1.2)	PC aa C36:4	0.357	0.104	5.07×10^{-13}	1.07×10^{-8}	-0.436
PC aa C40:4	PC ae C36:2	0.400	0.195	5.59×10^{-13}	1.07×10^{-8}	0.436
Sum(PC ae Cx:1.2)	Sum(PC aa Cx:4.5)	0.172	0.049	6.46×10^{-13}	1.07×10^{-8}	-0.434
PC aa C36:4	PC ae C36:2	16.802	6.170	6.94×10^{-13}	1.07×10^{-8}	0.434
PC aa C40:4	SM (OH) C20:1	0.657	0.436	7.08×10^{-13}	1.07×10^{-8}	0.434







