Research Repository UCD

Provided by the author(s) and University College Dublin Library in accordance with publisher policies. Please cite the published version when available.

Title	Metabolomic-based identification of clusters that reflect dietary patterns
Authors(s)	Gibbons, Helena; Carr, Eibhlin; McNulty, Breige A.; Walton, Janette; Flynn, Albert; Gibney, Michael J.; Brennan, Lorraine
Publication date	2017-07-20
Publication information	Molecular Nutrition and Food Research, 61 (10):
Publisher	Wiley
Item record/more information	http://hdl.handle.net/10197/8758
Publisher's statement	This is the author's version of the following article: Gibbons, Helena, Carr, Eibhlin, McNulty, Breige A., Walton, Janette, Flynn, Albert, Gibney, Michael J., Brennan, Lorraine : Metabolomic-based identification of clusters that reflect dietary patterns. Molecular Nutrition and Food Research, 2017, Which has been published in final form at https://doi.org/10.1002/mnfr.201601050
Publisher's version (DOI)	10.1002/mnfr.201601050

Downloaded 2022-01-27T11:58:22Z

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)

© Some rights reserved. For more information, please see the item record link above.

Metabolomic based identification of clusters that reflect dietary patterns

Helena Gibbons[1#], Eibhlin Carr[1#], Breige A McNulty[1], Anne P Nugent[1], Janette Walton[2], Albert Flynn[2], Michael J Gibney[1] and Lorraine Brennan[1].

[1]Institute of Food and Health, UCD School of Agriculture and Food Science, University College Dublin, Dublin, Republic of Ireland

[2]School of Food and Nutritional Sciences, University College Cork, Cork, Republic of Ireland

# both authors contributed equally

Address correspondence to Prof. Lorraine Brennan, Institute of Food and Health, Belfield, University College Dublin, Dublin 4, Ireland. E-mail: Lorraine.brennan@ucd.ie.

Abbreviations: CVD: cardiovascular disease, GI: glycemic index, IUNA: Irish Universities Nutrition Alliance, NANS: National Adult Nutritional Survey, PCA: Principal component analysis, PLS-DA: Partial least square discriminant analysis, VIP: Variable importance of projection.

Keywords: dietary patterns, dietary assessment, cluster analysis, metabolomics, nutritypes

Abstract (no more than 200 words)

Scope: Classification of subjects into dietary patterns generally relies on self-reporting dietary data which are prone to error. The aim of the present study was to develop a model for objective classification of people into dietary patterns based on metabolomic data.

Methods and results: Dietary and urinary metabolomic data from the National Adult Nutrition Survey (NANS) was used in the analysis (n=567). Two-step cluster analysis was applied to the urinary data to identify clusters. The subsequent model was used in an independent cohort to classify people into dietary patterns. Two distinct dietary patterns were identified. Cluster 1 was characterized by significantly higher intakes of breakfast cereals, low fat and skimmed milks, potatoes, fruit and fish, fish dishes ($P<0.05$) representing a "healthy" cluster. Cluster 2 had significantly higher intakes of chips/processed potatoes, meat products, savory snacks and high-energy beverages ($P<0.05$) representing an "unhealthy cluster". Classification was supported by significant differences in nutrient status ($P<0.05$). Validation in an independent group revealed that 94% of subjects were correctly classified.

Conclusion: The model developed was capable of classifying individuals into dietary patterns based on metabolomics data. Future applications of this approach could be developed for rapid and objective assignment of subjects into dietary patterns.

1 Introduction

Environmental factors such as diet play a major role in disease development. Traditionally, associations of single nutrients with disease risk have been studied for example increased saturated fat intake or vitamin D deficiency being associated with increased incident of cardiovascular disease (CVD) [1, 2]. However, individuals consume complex combinations of foods in different patterns with macronutrient, micronutrient and non-nutrient component interactions, therefore this single nutrient approach can fail to account for the complex interactions between nutrients within a meal [3, 4]. Therefore it is often difficult to separate out the specific effects of nutrients and the search for associations between single nutrients and disease may be problematic and confusing [5, 6]. In more recent years, classification of subjects into dietary patterns has emerged as a useful tool in nutritional epidemiology [7, 8]. The shift in focus to dietary patterns facilitates a more comprehensive relationship between nutrition and disease risk and also addresses issues of collinearity of nutrient intake and nutrient interactions [3]. Numerous studies have reported the beneficial effects of the DASH and Mediterranean diets on reducing blood pressure, preventing the risk of hypertension and reducing mortality in individuals with CVD [9-12], cancer [13, 14] and features of metabolic syndrome [15, 16]. The two most common types of dietary patterns that emerge from the literature are a healthy/prudent dietary pattern and an unhealthy/western dietary pattern, the healthy/prudent diet being associated with a reduced risk or lower incidence of disease compared to the western/unhealthy dietary patterns [17-20].

Approaches used to derive patterns from dietary data are based on a-priori criteria or a-posteriori statistical method. The a-priori approach is based on the use of dietary indices that aim to capture pre-defined healthy patterns, for example the Mediterranean Diet Score, designed to estimate adherence to the Greek variant of the Mediterranean diet [21, 22].

However, this approach focuses on specific features of diet, not accounting for the correlation structure of food and nutrient intakes and consequently, such scores do not reflect the overall effect of diet in general [23]. In contrast, for the a-posteriori multivariate statistical analysis (for example principal component analysis (PCA) and cluster analysis) is applied to the collected dietary information grouping individuals, within a population, who have similar dietary intake patterns. This approach ignores prior knowledge and takes into account many aspects of the diet rather than focusing on a few hypothesized key food groups [24]. However, both approaches rely on traditional methods of dietary assessment which are inherent with reporting bias and therefore may not provide an accurate estimate of food intake [25]. Biofluids such as saliva, urine and blood are potential rich sources of unbiased information about dietary behavior and the metabolic state of an individual. Therefore, in recent years, dietary biomarkers have been suggested as an unbiased objective measure of dietary intake [26, 27].

The application of metabolomics has emerged as a useful tool in the identification of biomarkers, however, to date, the use of metabolomic techniques in identifying biomarkers has focused largely on single foods [28-30]. An alternative approach has emerged whereby patterns of dietary intake are related to metabolomic profiles and the subsequent emergence of the term nutritype [31]. Nutritype refers to the classification of an individual into a dietary pattern based on their metabolomic profile. Assessment of metabolomic profiles and dietary intake has previously been described where urinary metabolomic profiles could be differentiated and characterized with respect to dietary group intake [31-34]. Although these studies provide good evidence for the link between habitual dietary patterns and metabolomic profiles, self-reporting dietary data is initially used to classify participants into the dietary patterns. Ideally one would like to be able to perform dietary pattern classification without

the need for self-reported data. Therefore, the objective of the present study was to develop a model that allows classification of individuals into a dietary pattern based on metabolomic profiles.

2 Materials and Methods

2.1 Sample population; identification of metabolomic profiles

Data from the National Adult Nutritional Survey (NANS) was used in this analysis. A detailed description of the methodology used in the NANS has been reported elsewhere [35, 36]. A brief outline is described; 1500 free-living adults aged 18 and over, representative of adults living in the Republic of Ireland with respect to age, gender, social class and urban/rural location were selected to participate in the survey. All eligible participants gave written consent according to the Helsinki declaration. For the present study 600 subjects were randomly selected ensuring equal number of males and females. Thirty three participants were excluded due to high ethanol and acetaminophen peaks therefore data was available for 567 participants. Ethical approval was obtained from University College Cork Clinical Research Ethics Committee of the Cork Teaching Hospitals (ECM 3 (p) 4 September 2008) and recruitment began in May 2008.

2.2 Dietary data collection, biofluid collection and anthropometry

Dietary data for the NANS participants was collected using four day semi-weighed food diaries. Participants were provided with digital scales and asked to weigh and record as many food items as possible that were consumed. Information regarding the cooking methods, brand names and recipes was also recorded. To guarantee accurate recording of dietary intake, a researcher visited participants in their homes or workplaces three times during the recording period. A quantification protocol established by the Irish Universities Nutrition

Alliance (IUNA) for a previous survey [37] was updated for the NANS. Food intake data were analyzed using WISP© (Tinuviel Software, Anglesey, UK) which uses data from, McCance and Widdowsons's, The Composition of Foods, the fifth and sixth editions plus all nine supplemental volumes to generate nutrient intake. The food data was subsequently reduced into 68 food groups and for the purpose of this analysis was further aggregated into 33 food groups as previously defined [38] . Food groups were expressed as a percentage of energy intake using the mean proportion at the individual level [39]. When the dietary recording period had ended, participants provided a 50 ml first void urine sample. Samples were kept on ice during transport to the lab for processing and stored at -80ºC for analysis. Participants in the present study also provided a fasting blood sample. Collection and processing of blood samples has previously been described [38, 40].

Anthropometric measurements were also taken by the researcher in the participant's home and included weight (kg), height (m) and body composition. Height was measured with the head positioned in the Frankfurt Plane. Weight and body composition were measured using a Tanita body composition analyzer BC-420MA (Tanita Ltd, GB). All measurements were performed in duplicate. Blood pressure was also measured by the researcher in triplicate using the OMRON M6 comfort blood pressure monitor.

2.3 Validation study; The NutriTech food intake study

Participants from the NutriTech food intake study were used to investigate the ability of the model to classify people into different dietary patterns. The NutriTech food intake study aimed to investigate the use of metabolomic profiling as a method of independent food quantification. This study involved 50 participants being randomly assigned to one of five different diet groups; red meat, fish, poultry, processed meat or a supplement and vegetarian

option (Supporting Information Table S1). Eligibility criteria included males and females of all ethnicities, aged between 18 and 65 y with a BMI of 18.5-35 kg/m^2 and free from any chronic medical condition (study characteristics are described in Supporting Information Table S2). Ethical approval was received from London Brent Ethics Committee (reference number: 12/LO/0139) and written informed consent was provided by all participants. Participants attended the NIHR/Wellcome Trust Imperial Clinical Research Facility for three days over three consecutive weeks. During this time participants consumed a standardized breakfast at 8 am and their test meals at midday (12 pm) and evening (7 pm). All test meals were designed to provide similar intakes of dietary energy and fiber but macronutrient composition varied over the intervention weeks with carbohydrate decreasing from week 1 to week 3 and protein and fat intake increasing from week 1 to week 3. Fasting spot and postprandial urine and plasma samples were collected. The week 2 fasting spot urine samples were used in this analysis as the week 2 diet consisted of the recommended macronutrient distribution for a healthy diet (20%, 35% and 45% of total energy intake from protein, fat, and carbohydrate respectively). Furthermore, a number of unhealthy foods and beverages were also absent from the participant's diet (Supporting Information Table S3).

2.4 Urine analysis

Urine samples (500 µL) were prepared by the addition of 250 µL potassium phosphate buffer (0.2 M KH_2PO_4, 0.8 M K_2HPO_4). After centrifugation at 5360 g for 5 min, 50 µL deuterium oxide and 10 µL sodium trimethyl [2,2,3,3-2H4] propionate (TSP) (0.005 g/mL) were added to 540 µL of the supernatant sample. ^1H NMR spectra were acquired on a 600-MHz Varian NMR spectrometer (Varian Limited, Oxford, United Kingdom) by using the first increment of a NOESY (Nuclear overhauser effect spectroscopy) pulse (sequence at 25°C). Spectra were acquired with 16384 data points and 128 scans. Water suppression was achieved during

the relaxation delay (2.5 s) and the mixing time (100ms). All spectra were referenced to TSP at 0.0 ppm. ^{1}H NMR spectra were processed manually with Chenomx software (version 7.5; Chenomx Edmonton, Canada) by using a line broadening of 0.2 Hz. and were phase and baseline corrected. Spectra were converted into 550 spectral regions of 0.01 ppm. Spectral regions from 0.505 ppm to 7.995 ppm inclusive were analyzed. The water region was excluded, and data were normalized to the total area of the spectral integral. Discriminating metabolites were identified by using libraries of pure metabolites in house and the Chenomx database library.

2.5 Biochemical analysis

Biochemistry values were assessed using a clinical bioanalyzer (RX Daytona; Randox Laboratories). Details of analytical methods for serum triglycerides, total cholesterol, HDL cholesterol, glucose, non-esterified fatty acids, C - reactive protein, insulin, creatinine, potassium, sodium, 25-hydroxyvitamin D and urinary sodium, potassium and creatinine have previously been reported [40]. Serum LDL cholesterol was calculated using the Friedewald equation [41]. Red cell folate which is reflective of longer term folate exposure and serum folate which is reflective of short term exposure were measured using a microbiological assay. Serum vitamin B-12 was also measured by microbiological assay and total plasma homocysteine (tHcy) was measured by fluorescence polarization immunoassay. Hemoglobin was measured using the Beckman Coulter AcT diff™ analyzer. Long term riboflavin status was determined by erythrocyte glutathione reductase activation coefficient (EGRac) functional assay using an autoanalyzer. A reverse-phase high performance liquid chromatograph (HPLC) method was used in the measurement of plasma pyridoxal-5'-phosphate (PLP). All samples were run in duplicate and standard quality control procedures were followed on all analyzers to ensure data integrity.

2.6 Statistical Analysis

The NANS urinary metabolomic data was standardized by subtracting the variables minimum value from each value and dividing by the variables range as previously described [42, 43]. Two-step cluster analysis was employed (IBM SPSS version 20.0) on the NMR urine data to define the number of clusters using the Schwarz's Bayesian inference criterion (BIC). This analysis included age and gender variables in the data matrix. Multivariate statistical analysis was performed using Simca-P+ software (version 13.0; Umetrics). Prior to data analyses the NANS and NutriTech datasets were scaled using pareto scaling [44]. PCA and PLS-DA was applied to the NANS data set to explore any trends in the data. Orthogonal partial least squares analysis (OPLS-DA) which improves interpretation and separation between classes on a scores plot by filtering unwanted variation was also applied to this data set [45]. Permutation testing, using 50 random permutations was used to confirm model validity. To identify the discriminating variables responsible for the class membership, the variable importance in the projection (VIP) value of each variable was calculated. A variable with a VIP close to or greater than 2 can be considered important in a given model[46].

To investigate the robustness of the PLS-DA model, a training set and test set were constructed. Two-thirds of the NANS data ($n = 378$) were defined as the training set and the remaining NANS samples were defined as the test set ($n = 189$). The model's ability to correctly classify or predict the test set into its original classes was calculated.

Statistical analyses were performed using IBM SPSS (version 20.0). Data were assessed for normality and where appropriate non parametric tests or log transformation was used. Independent t-tests were used to indicate whether a difference in metabolites was statistically

significant between the clusters, P-values were corrected for multiple comparisons using a Bonferroni adjustment. Chi-square analysis was used to explore the distribution of males and females and supplement users between clusters.

3 Results

3.1 Identification of clusters using urinary metabolomic data from the NANS cohort

567 subjects were included in the study, 286 males and 281 females. Using two-step cluster analysis a total of two clusters were found in the NMR metabolomics data. Subject characteristics of both clusters are presented in Table 1. Cluster 1 had a significantly higher mean BMI and age compared to cluster 2 (P = 0.043 and 1.5×10^{-14} respectively). Cluster 1 also had a significantly lower energy intake compared to cluster 2 (P = 0.025). There was a significant difference in supplement users across the clusters (P = 0.001) with a higher proportion of supplement users in cluster 1. There was a small non-significant difference in gender distribution between the two clusters.

PCA was used for visualization of the predetermined clusters (Figure 1A). A robust PLS-DA model (Figure 1B) was obtained (R^2X: 0.145, Q^2: 0.578, Q^2 intercept for permutation testing 0.0, -0.13). To further examine differences in metabolomic profiles between the clusters, a VIP list was obtained from the PLS-DA model and a metabolite was assigned to each peak. Cluster 1 had higher levels of hippurate, betaine, anserine, N-phenylacetylglutamine, 3-hydroxybutyrate, citrate, tryptophan and 2-aminoadipate, cluster 2 had higher levels of creatinine, glycylproline, N-acetylglutamate and theophylline (Table 2).

Investigation of the robustness of the PLS-DA model demonstrated that a high percentage (Cluster 1: 91.9%, Cluster 2: 90.0%) of subjects from the test set were correctly predicted into the training set they were originally assigned to (Supporting Information Table S4).

3.2 Dietary intake, nutrient profiles and biochemical profiles are significantly different across the clusters

Analysis of dietary intake across the cluster groups revealed significant differences (Table 3). Two distinct food intake patterns were identified; cluster 1 was characterized by significantly higher intakes of breakfast cereals and porridge, low fats and skimmed milks, potatoes, fruit, fish, fish dishes and fish products ($P < 0.05$). Cluster 2 had significantly higher intakes of chips and processed potatoes, meat products, savory snacks and high energy beverages ($P < 0.05$). Nutrient profiles, obtained from the dietary data, were also compared between the two clusters (Table 4). Cluster 1 had significantly higher intakes of carbohydrates, protein, sugars and dietary fiber, and overall had a significantly more nutrient dense profile compared to cluster 2 ($P < 0.05$). Comparison of these patterns resulted in cluster 1 being defined the "healthy" cluster as it was characterized by a higher mean percentage energy intake of the more nutritionally favorable food groups. Cluster 2 was defined the "unhealthy" cluster as it was characterized by a higher mean percentage intake of the more energy dense and nutritionally undesirable food groups including meat products, savory snacks and high energy beverages.

Comparison of the biochemical profiles and biomarkers of nutrient status across the clusters revealed a number of significant differences that were reflective of the dietary cluster classification (Table 5). Cluster 1 had significantly higher HDL cholesterol, serum folate and red cell folate ($P < 0.05$), while cluster 2 had significantly higher EGRac ($P < 0.05$). Urinary sodium and creatinine were also significantly higher in cluster 2.

3.3 Cluster validation using data from the NutriTech food intake study

The validity of the cluster model was investigated using urinary metabolomics data from an independent group, the NutriTech food intake study participants. These participants underwent an intervention in a controlled environment and received a diet consisting of the recommended macronutrient distribution for a healthy diet. Metabolomic data was available for 49 participants (22 males and 27 females). Using the model built from the NANS data, cluster membership of the NutriTech food intake participants was predicted. As these participants were following a healthy diet in a controlled environment we expected that they would be classified into the healthy dietary cluster. Importantly, the analysis revealed that 94% of the participants were correctly classified into the healthy cluster.

4 Discussion

The present study has advanced the classification of individuals into dietary patterns through the use of urinary metabolomic data. The classification was supported by significant differences between the patterns in nutrient intakes and biochemical profiles. Furthermore, application to an independent study demonstrated the utility of the approach for rapid classification of subjects into dietary patterns.

The novel aspect of this work is the development of a model, using urinary metabolomics data, which can be used to classify individuals into dietary patterns. This therefore eliminates the reliance on self-reporting dietary data which is inherently inaccurate [25, 47, 48]. This work is an important demonstration of the classification of individuals into dietary patterns using metabolomics data only. Previous literature linking dietary patterns to metabolomic profiles have used self-reported dietary data to identify dietary patterns and have examined

only associations or relationships between the metabolomics data and dietary patterns [31, 49-52]. Furthermore, the dietary patterns identified within this research were also corroborated by significant differences in nutrient status. The model was further validated using an independent cohort; the NutriTech food intake study.

Dietary patterns across the two clusters revealed significant differences in intakes of certain food groups and nutrient profiles. Cluster 1, characterized as a healthy dietary pattern and cluster 2 as an unhealthy dietary pattern correspond closely with previous dietary pattern analyses, where opposing dietary patterns, "prudent" (higher consumption of nutritionally favorable foods) and "western" (higher consumption of nutritionally unfavorable food) tend to emerge from population groups [53, 54]. In a sample of Irish adolescents five dietary clusters, including a healthy and unhealthy dietary pattern were identified [55]. A previous study analyzing a middle-aged Irish population identified three dietary clusters as follows; traditional Irish diet (highest intakes of white bread, butter, whole milk, dairy products desserts and sweets), a prudent diet (higher intakes of brown bread, unrefined cereals, fish, low fat dairy products, fruit, vegetables and poultry) and an unhealthy diet termed "alcohol and convenience food diet" (higher intakes of alcohol, meat, meat products and chips) [56]. A more recent study identified three dietary patterns, again revealing clusters characterized by traditional, prudent and unhealthy diets [31]. The dietary patterns in the present study correspond closely with those previously published with the traditional dietary patterns containing elements of both the healthy and unhealthy dietary pattern.

Biochemical profiles reflected and therefore supported the classification of the dietary clusters. Cluster 1 had significantly higher serum folate and red cell folate while cluster 2 had significantly higher EGRac concentrations. A higher EGRac result corresponded to the lower

intake of low fat and skimmed milks, breakfast cereals, fruit, vegetables, fish and yoghurts. The significantly higher serum folate and red cell folate in cluster 1 was reflective of the higher levels of the nutrient folate and corresponded to the higher intake of folate rich foods such as breakfast cereals, wholemeal and brown breads, vegetables and potatoes. A higher urinary sodium may be associated with a higher intake of processed meats in cluster 2, which contributes to 18% of mean sodium intake in the Irish population [35] and this may also be associated with the higher intake of chips and processed potatoes. Overall, the nutrient status data support the dietary patterns obtained using the metabolomics data.

Examination of the urinary markers that contributed to the determination of the dietary clusters revealed that many have been previously linked with dietary intakes. Hippurate, N-phenylacetylglutamine, betaine, 3-Hydroxybutyrate, and anserine were associated with certain dietary components that are present in the current healthy cluster. For example, hippurate and N-phenylacetylglutamine have previously been associated with the consumption of plant based foods, vegetarian diets and green and black tea [57-60], and betaine with intakes of citrus fruits and legumes [28, 61]. Elevated concentrations of 3-Hydroxybutyrate has previously been reported after the consumption of acidified milk products [62] and anserine has been reported in poultry, meats and oily fish [63]. The metabolites found at higher levels in cluster 2, glycylproline, theophylline and N-acetylglutamate (NAG) have previously been associated with foods found in the current unhealthy pattern. Theophylline, an alkaloid caffeine analogue, is naturally found in coffee, chocolate, tea and as an additive in soft drinks [64-66] and NAG has been identified in coffee, tea and soybeans with highest concentrations in cocoa powder [67]. Glycylproline is an end product of collagen metabolism and the addition of collagen to foods such as meat products is common practice due to its water binding properties [68].

There are a number of strengths associated with the present study. Firstly the objective identification of dietary patterns using urinary metabolomics data and the use of biochemical profiles to support the cluster classification. Secondly, the model was successfully validated in an independent cohort therefore demonstrating the utility of this model for the rapid and objective assignment of subjects to a dietary pattern. This could further be developed for delivery of personalized nutrition, as previously described in relation to metabotypes, through the application of decision trees to derive dietary advice for the clusters identified [40]. However, it must be noted that the dietary patterns identified in this analysis are pre-defined and therefore may introduce the disadvantages of the a-priori approach. Furthermore demonstration of this approach in other diverse study populations will be imperative for future progression of this concept. Moreover, the stability and reproducibility of these dietary patterns may need to be assessed in longer-term studies.

In conclusion, this study successfully classified subjects into two distinct dietary patterns using metabolomics data. Importantly, this classification was supported by dietary data, nutrient status and validation in an independent cohort. This represents a novel approach to assessment of dietary intake and furthermore, represents a significant advancement in the potential use of dietary markers.

Author contributions

The authors' responsibilities were as follows—HG, EC: conducted research, analyzed data and wrote paper; BAM, APN, JW, AF and MJG: provided essential materials; LB: designed research, conducted research, analyzed data and wrote paper. All authors read and approved the final manuscript.

Acknowledgements/funding

We thank all volunteers for their commitment and patience during the study.

HG and EC are supported by a research grant from NutriTech (project number: 2896511). LB acknowledges the following funding: The European Research Council ERC (647783) and SFI (JPI-HDHL/B3075, Foodball). Data used in part of this analysis were supported by joint funding from the Irish Department of Agriculture, Fisheries and Food and the Health Research Board under the Food for Health Research Initiative (2007–2012) (grant number 7FHRIUCC2)

Conflict of interest statement

The authors have declared no conflict of interest.

5 References

[1] Wang, T. J., Pencina, M. J., Booth, S. L., Jacques, P. F., et al., Vitamin D Deficiency and Risk of Cardiovascular Disease. *Circulation.* 2008, *117*, 503-511.
[2] Xu, J., Eilat-Adar, S., Loria, C., Goldbourt, U., et al., Dietary fat intake and risk of coronary heart disease: the Strong Heart Study. *Am. J. Clin. Nutr.* 2006, *84*, 894-902.
[3] Jacques, P. F., Tucker, K. L., Are dietary patterns useful for understanding the role of diet in chronic disease? *Am. J. Clin. Nutr.* 2001, *73*, 1-2.
[4] Cespedes, E. M., Hu, F. B., Dietary patterns: from nutritional epidemiologic analysis to national guidelines. *Am. J. Clin. Nutr.* 2015, *101*, 899-900.
[5] Randall, E., Marshall, J. R., Graham, S., Brasure, J., Patterns in food use and their associations with nutrient intakes. *Am. J. Clin. Nutr.* 1990, *52*, 739-745.
[6] Gerber, M., The Comprehensive Approach to Diet: A Critical Review. *J. Nutr.* 2001, *131*, 3051S-3055S.
[7] Fung, T. T., Rimm, E. B., Spiegelman, D., Rifai, N., et al., Association between dietary patterns and plasma biomarkers of obesity and cardiovascular disease risk. *Am. J. Clin. Nutr.* 2001, *73*, 61-67.
[8] Huijbregts, P., Feskens, E., Kromhout, D., Dietary Patterns and Cardiovascular Risk Factors in Elderly Men: The Zutphen Elderly Study. *Int. J. Epidemiol.* 1995, *24*, 313-320.
[9] Lopez-Garcia, E., Rodriguez-Artalejo, F., Li, T. Y., Fung, T. T., et al., The Mediterranean-style dietary pattern and mortality among men and women with cardiovascular disease. *Am. J. Clin. Nutr.* 2014, *99*, 172-180.
[10] Rees, K., Hartley, L., Flowers, N., Clarke, A., et al., 'Mediterranean' dietary pattern for the primary prevention of cardiovascular disease. Cochrane. Database. Syst. Rev. 2013, 12, 1-53.
[11] Sacks, F. M., Appel, L. J., Moore, T. J., Obarzanek, E., et al., A dietary approach to prevent hypertension: a review of the Dietary Approaches to Stop Hypertension (DASH) Study. *Clin. Cardiol.* 1999, *22*, 1116-1120.
[12] Sacks, F. M., Svetkey, L. P., Vollmer, W. M., Appel, L. J., et al., Effects on Blood Pressure of Reduced Dietary Sodium and the Dietary Approaches to Stop Hypertension (DASH) Diet. *N. Engl. J. Med.* 2001, *344*, 3-10.
[13] Fung, T. T., Hu, F. B., Wu, K., Chiuve, S. E., et al., The Mediterranean and Dietary Approaches to Stop Hypertension (DASH) diets and colorectal cancer. *Am. J. Clin. Nutr.* 2010, *92*, 1429-1435.
[14] Couto, E., Boffetta, P., Lagiou, P., Ferrari, P., et al., Mediterranean dietary pattern and cancer risk in the EPIC cohort. *Br. J. Cancer.* 2011, *104*, 1493-1499.
[15] Azadbakht, L., Mirmiran, P., Esmaillzadeh, A., Azizi, T., et al., Beneficial Effects of a Dietary Approaches to Stop Hypertension Eating Plan on Features of the Metabolic Syndrome. *Diabetes. Care.* 2005, *28*, 2823-2831.
[16] Kastorini, C. M., Milionis, H. J., Esposito, K., Giugliano, D., et al., The Effect of Mediterranean Diet on Metabolic Syndrome and its Components: A Meta-Analysis of 50 Studies and 534,906 Individuals. *J. Am. Coll. Cardiol.* 2011, *57*, 1299-1313.
[17] Bertuccio, P., Rosato, V., Andreano, A., Ferraroni, M., et al., Dietary patterns and gastric cancer risk: a systematic review and meta-analysis. *Ann. Oncol.* 2013, *24*, 1450-1458.
[18] Brennan, S. F., Cantwell, M. M., Cardwell, C. R., Velentzis, L. S., et al., Dietary patterns and breast cancer risk: a systematic review and meta-analysis. *Am. J. Clin. Nutr.* 2010, *91*, 1294-1302.
[19] Fung, T. T., Schulze, M., Manson, J. E., Willett, W. C., et al., Dietary patterns, meat intake, and the risk of type 2 diabetes in women. *Arch. Intern. Med.* 2004, *164*, 2235-2240.
[20] Stricker, M. D., Onland-Moret, N. C., Boer, J. M. A., van der Schouw, Y. T., et al., Dietary patterns derived from principal component- and k-means cluster analysis: Long-term association with coronary heart disease and stroke. *Nutr. Metab. Cardiovasc. Dis.* 2013, *23*, 250-256.
[21] Kastorini, C. M., Papadakis, G., Milionis, H. J., Kalantzi, K., et al., Comparative analysis of a-priori and a-posteriori dietary patterns using state-of-the-art classification algorithms: A case/case-control study. *Artif. Intell. Med.* 2013, *59*, 175-183.

[22] Trichopoulou, A., Costacou, T., Bamia, C., Trichopoulos, D., Adherence to a Mediterranean Diet and Survival in a Greek Population. *N. Engl. J. Med.* 2003, *348*, 2599-2608.
[23] Hoffmann, K., Schulze, M. B., Schienkiewitz, A., Nöthlings, U., et al., Application of a New Statistical Method to Derive Dietary Patterns in Nutritional Epidemiology. *Am. J. Epidemiol.* 2004, *159*, 935-944.
[24] Nettleton, J. A., Schulze, M. B., Jiang, R., Jenny, N. S., et al., A priori–defined dietary patterns and markers of cardiovascular disease risk in the Multi-Ethnic Study of Atherosclerosis (MESA). *Am. J. Clin. Nutr.* 2008, *88*, 185-194.
[25] Kipnis, V., Midthune, D., Freedman, L., Bingham, S., et al., Bias in dietary-report instruments and its implications for nutritional epidemiology. *Public. Health. Nutr.* 2002, *5*, 915-923.
[26] Potischman, N., Biologic and methodologic issues for nutritional biomarkers. *J. Nutr.* 2003, 133 Suppl *3*, 875s-880s.
[27] Freedman, L. S., Kipnis, V., Schatzkin, A., Tasevska, N., et al., Can we use biomarkers in combination with self-reports to strengthen the analysis of nutritional epidemiologic studies? *Epidemiol. Perspect. Innov.* 2010, *7*, 2.
[28] Heinzmann, S. S., Brown, I. J., Chan, Q., Bictash, M., et al., Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. *Am. J. Clin. Nutr.* 2010, *92*, 436-443.
[29] Cross, A. J., Major, J. M., Sinha, R., Urinary Biomarkers of Meat Consumption. Cancer. Epidemiol. Biomarkers. Prev*.* 2011, *20*, 1107-1111.
[30] Altorf-van der Kuil, W., Brink, E. J., Boetje, M., Siebelink, E., et al., Identification of biomarkers for intake of protein from meat, dairy products and grains: a controlled dietary intervention study. *Br. J. Nutr.* 2013, *110*, 810-822.
[31] O'Sullivan, A., Gibney, M. J., Brennan, L., Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. *Am. J. Clin. Nutr.* 2011, *93*, 314-321.
[32] Rasmussen, L. G., Winning, H., Savorani, F., Toft, H., et al., Assessment of the Effect of High or Low Protein Diet on the Human Urine Metabolome as Measured by NMR. *Nutrients.* 2012, *4*, 112-131.
[33] Rasmussen, L. G., Winning, H., Savorani, F., Ritz, C., et al., Assessment of dietary exposure related to dietary GI and fibre intake in a nutritional metabolomic study of human urine. *Genes. Nutr.* 2012, *7*, 281-293.
[34] Andersen, M. S., Rinnan, Å., Manach, C., Poulsen, S. K., et al., Untargeted Metabolomics as a Screening Tool for Estimating Compliance to a Dietary Pattern. *J. Proteome. Res.* 2014, *13*, 1405-1418.
[35] Irish Universities Nutrition Alliance. National Adult Nutrition Survey Summary Report. 2011.
[36] Cashman, K. D., Muldowney, S., McNulty, B., Nugent, A., et al., Vitamin D status of Irish adults: findings from the National Adult Nutrition Survey. *Br. J. Nutr.* 2013, *109*, 1248-1256.
[37] Harrington, K. E., Robson, P. J., Kiely, M., Livingstone, M. B., et al., The North/South Ireland Food Consumption Survey: survey design and methodology. *Public. Health. Nutr.* 2001, *4*, 1037-1042.
[38] Li, K., Brennan, L., McNulty, B. A., Bloomfield, J. F., et al., Plasma fatty acid patterns reflect dietary habits and metabolic health: A cross-sectional study. *Mol. Nutr. Food. Res.* 2016, *60*, 2043-2052.
[39] Krebs-Smith, S. M., Kott, P. S., Guenther, P. M., Mean proportion and population proportion two answers to the same question. *J. Am. Diet. Assoc.* 1989, *89*, 667-672.
[40] O'Donovan, C. B., Walsh, M. C., Nugent, A. P., McNulty, B., et al., Use of metabotyping for the delivery of personalised nutrition. *Mol. Nutr. Food. Res.* 2015, *59*, 377-385.
[41] Friedewald, W. T., Levy, R. I., Fredrickson, D. S., Estimation of the Concentration of Low-Density Lipoprotein Cholesterol in Plasma, Without Use of the Preparative Ultracentrifuge. *Clin. Chem.* 1972, *18*, 499-502.

[42] Lo Siou, G., Yasui, Y., Csizmadi, I., McGregor, S. E., et al., Exploring Statistical Approaches to Diminish Subjectivity of Cluster Analysis to Derive Dietary Patterns: The Tomorrow Project. *Am. J. Epidemiol.* 2011, *173*, 956-967.
[43] Milligan, G. W., Cooper, M. C., A study of standardization of variables in cluster analysis. *J. Classif.* 1988, *5*, 181-204.
[44] Worley, B., Powers, R., Multivariate Analysis in Metabolomics. *Curr. Metabolomics.* 2013, *1*, 92-107.
[45] Trygg, J., Wold, S., Orthogonal projections to latent structures (O-PLS). *J. Chemom.* 2002, *16*, 119-128.
[46] L. Eriksson, E. J., N. Kettaneh-Wold, J. Trygg, C. Wikström and S. Wold, Multi- and megavariate data analysis. Umetrics AB. 2006.
[47] Bingham, S. A., Biomarkers in nutritional epidemiology. *Public. Health. Nutr.* 2002, *5*, 821-827.
[48] Dhurandhar, N. V., Schoeller, D., Brown, A. W., Heymsfield, S. B., et al., Energy balance measurement: when something is not better than nothing. *Int. J. Obes.* 2015, *39*, 1109-1113.
[49] Altmaier, E., Kastenmüller, G., Römisch-Margl, W., Thorand, B., 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.
[50] Floegel, A., von Ruesten, A., Drogan, D., Schulze, M. B., et al., Variation of serum metabolites related to habitual diet: a targeted metabolomic approach in EPIC-Potsdam. *Eur. J. Clin. Nutr.* 2013, *67*, 1100-1108.
[51] Peré-Trepat, E., Ross, A. B., Martin, F.-P., Rezzi, S., et al., Chemometric strategies to assess metabonomic imprinting of food habits in epidemiological studies. *Chemometr. Intell. Lab.* 2010, *104*, 95-100.
[52] Bouchard-Mercier, A., Rudkowska, I., Lemieux, S., Couture, P., et al., The metabolic signature associated with the Western dietary pattern: a cross-sectional study. *Nutr. J.* 2013, *12*, 1-9.
[53] Maghsoudi, Z., Azadbakht, L., How dietary patterns could have a role in prevention, progression, or management of diabetes mellitus? Review on the current evidence. *J. Res. Med. Sci.* 2012, *17*, 694-709.
[54] Lopez-Garcia, E., Schulze, M. B., Fung, T. T., Meigs, J. B., et al., Major dietary patterns are related to plasma concentrations of markers of inflammation and endothelial dysfunction. *Am. J. Clin. Nutr.* 2004, *80*, 1029-1035.
[55] Hearty, Á. P., Gibney, M. J., Dietary patterns in Irish adolescents: a comparison of cluster and principal component analyses. *Public. Health. Nutr.* 2013, *16*, 848-857.
[56] Villegas, R., Salim, A., Collins, M. M., Flynn, A., et al., Dietary patterns in middle-aged Irish men and women defined by cluster analysis. *Public. Health. Nutr.* 2004, *7*, 1017-1024.
[57] Xu, J., Yang, S., Cai, S., Dong, J., et al., Identification of biochemical changes in lactovegetarian urine using 1H NMR spectroscopy and pattern recognition. *Anal. Bioanal. Chem.* 2010, *396*, 1451-1463
[58] Krupp, D., Doberstein, N., Shi, L., Remer, T., Hippuric acid in 24-hour urine collections is a potential biomarker for fruit and vegetable consumption in healthy children and adolescents. *J. Nutr.* 2012, *142*, 1314-1320.
[59] Walsh, M. C., Brennan, L., Pujos-Guillot, E., Sébédio, J.-L., et al., Influence of acute phytochemical intake on human urinary metabolomic profiles. *Am. J. Clin. Nutr.* 2007, *86*, 1687-1693.
[60] Mulder, T. P., Rietveld, A. G., van Amelsvoort, J. M., Consumption of both black tea and green tea results in an increase in the excretion of hippuric acid into urine. *Am. J. Clin. Nutr.* 2005, *81*, 256S-260S.
[61] Atkinson, W., Downer, P., Lever, M., Chambers, S. T., et al., Effects of orange juice and proline betaine on glycine betaine and homocysteine in healthy male subjects. *Eur. J. Nutr.* 2007, *46*, 446-452.

[62] Pedersen, S. M., Nielsen, N. C., Andersen, H. J., Olsson, J., et al., The serum metabolite response to diet intervention with probiotic acidified milk in irritable bowel syndrome patients is indistinguishable from that of non-probiotic acidified milk by 1H NMR-based metabonomic analysis. *Nutrients.* 2010, *2*, 1141-1155.
[63] Kubomura, D., Matahira, Y., Masui, A., Matsuda, H., Intestinal Absorption and Blood Clearance of l-Histidine-Related Compounds after Ingestion of Anserine in Humans and Comparison to Anserine-Containing Diets. *J. Agric. Food. Chem.* 2009, *57*, 1781-1785.
[64] Iriti, M., Varoni, E. M., Vitalini, S., Melatonin in traditional Mediterranean diets. *J. Pineal. Res.* 2010, *49*, 101-105.
[65] Bispo, M. S., Veloso, M. C. C., Pinheiro, H. L. C., De Oliveira, R. F. S., et al., Simultaneous Determination of Caffeine, Theobromine, and Theophylline by High-Performance Liquid Chromatography. *J. Chromatogr. Sci.* 2002, *40*, 45-48.
[66] Chen, Q., Wang, J., Simultaneous determination of artificial sweeteners, preservatives, caffeine, theobromine and theophylline in food and pharmaceutical preparations by ion chromatography. *J. Chromatogr. A.* 2001, *937*, 57-64.
[67] Hession, A. O., Esrey, E. G., Croes, R. A., Maxwell, C. A., N-Acetylglutamate and N-Acetylaspartate in Soybeans (Glycine max L.), Maize (Zea maize L.), and Other Foodstuffs. *J. Agric. Food. Chem.* 2008, *56*, 9121-9126.
[68] Mazorra-Manzano, M., Torres-Llanez, M., González-Córdova, A., Vallejo-Cordoba, B., A Capillary Electrophoresis Method for the Determination of Hydroxyproline as a Collagen Content Index in Meat Products. *Food. Analytical. Methods.* 2012, *5*, 464-470.

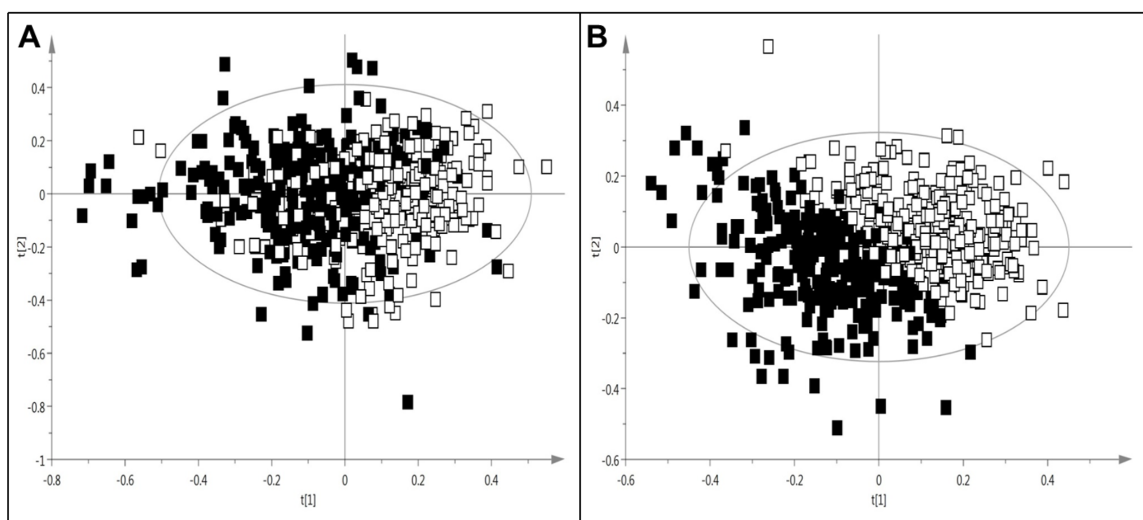

FIGURE 1

Figure 1 A Principal components analysis (PCA) and Figure 1 B Partial least squares (PLS-DA) of ^{1}H NMR urine data of clusters

Clusters were determined from metabolite concentrations by using two-step cluster analysis. Cluster 1 (■), Cluster 2 (□), **A** PCA of clusters. R^2X: 0.177 and Q^2:0.0703. **B** PLS-DA of clusters 1 and 2. R^2X: 0.145 and Q^2:0.578.

Table 1 Subject characteristics across the two clusters[a)]

Characteristic	Cluster 1 (n=286)	Cluster 2 (n=281)	P[b)]
Gender	133 M 153 F	153 M 128 F	0.059
Energy intake (MJ)	8.3 ± 2.5	**8.8 ± 2.6**	0.025
BMI (kg/m^2)	**31.1 ± 16.3**	28.7 ± 11.1	0.043
Age (y)	**52 ± 16**	42 ± 15	1.5x10^{-14}
Supplement Users (%)	**42**	29	0.001
Systolic BP (mm/Hg)	126.6 ± 18.2	125.2 ± 18.7	0.391
Diastolic BP (mm/Hg)	79.1 ± 10.7	78.5 ± 11.4	0.566

[a)]Mean ± SD (all such values). Clusters were determined using two-step cluster analysis.

[b)] Significant differences in gender and supplement users based on Chi-square. Significant differences in energy, BMI, age, systolic and diastolic BP based on independent t-test. Bold values indicate significantly higher values. BP, Blood Pressure.

Table 2 Metabolite differences between the clusters [a)]

Peaks[b)] (ppm)	VIP[c)]	Cluster 1	Cluster 2	Metabolites	P[d)]
3.955	5.18	**1.59 ± 0.66**	1.15 ± 0.42	Hippurate	$6.3x10^{-18}$
3.035	5.02	9.19 ± 2.07	**10.00 ± 2.13**	Creatinine	$2.5x10^{-4}$
3.255	4.53	**1.35 ± 1.68**	0.79 ± 0.61	Betaine	$6.4x10^{-6}$
3.965	4.00	**1.18 ± 0.43**	0.91 ± 0.29	Hippurate	$2.8x10^{-16}$
7.815	3.99	**0.84 ± 0.47**	0.56 ± 0.30	Hippurate	$5.9x10^{-15}$
7.545	3.92	**0.71 ± 0.37**	0.47 ± 0.24	Hippurate	$5.3x10^{-18}$
7.345	3.38	**0.47 ± 0.21**	0.31 ± 0.11	N-phenylacetylglutamine	$2.8x10^{-24}$
7.825	3.38	**0.64 ± 0.38**	0.43 ± 0.24	Hippurate	$4.0x10^{-13}$
7.555	3.20	**0.51 ± 0.26**	0.34 ± 0.17	Hippurate	$5.5x10^{-17}$
7.355	3.07	**0.39 ± 0.16**	0.27 ± 0.09	N-phenylacetylglutamine	$3.3x10^{-26}$
7.835	2.78	**0.45 ± 0.23**	0.31 ± 0.16	Hippurate	$1.5x10^{-14}$
2.645	2.61	**0.68 ± 0.49**	0.48 ± 0.45	Anserine	$2.9x10^{-5}$
7.415	2.58	**0.30 ± 0.11**	0.21 ± 0.07	Hippurate	$6.8x10^{-27}$
7.335	2.58	**0.32 ± 0.11**	0.23 ± 0.08	N-phenylacetylglutamine	$9.7x10^{-26}$
2.335	2.51	**0.49 ± 0.20**	0.37 ± 0.16	3-hydroxybutyrate	$5.5x10^{-13}$
7.535	2.37	**0.38 ± 0.24**	0.26 ± 0.15	Hippurate	$2.1x10^{-10}$
2.255	2.34	**0.35 ± 0.11**	0.27 ± 0.07	2-aminoadipate	$1.4x10^{-21}$
2.535	2.31	**0.96 ± 0.43**	0.79 ± 0.48	Citrate	$3.1x10^{-4}$
7.625	2.27	**0.29 ± 0.17**	0.20 ± 0.11	Hippurate	$3.1x10^{-13}$
7.405	2.16	**0.24 ± 0.08**	0.17 ± 0.06	N-phenylacetylglutamine	$2.9x10^{-24}$
7.615	2.15	**0.25 ± 0.13**	0.16 ± 0.09	Hippurate	$2.9x10^{-15}$
3.045	2.06	0.53 ± 0.20	**0.65 ± 0.29**	Creatinine	$3.0x10^{-6}$

3.515	1.98	0.30 ± 0.06	**0.35 ± 0.08**	Glycylproline	$1.8x10^{-20}$
3.505	1.94	0.35 ± 0.08	**0.42 ± 0.12**	Glycylproline	$4.6x10^{-14}$
7.635	1.93	**0.19 ± 0.09**	0.13 ± 0.06	Hippurate	$4.9x10^{-17}$
3.565	1.92	0.34 ± 0.09	**0.41 ± 0.15**	Theophylline	$4.5x10^{-11}$
7.365	1.89	**0.19 ± 0.08**	0.13 ± 0.06	N-phenylacetylglutaime	$7.4x10^{-19}$
7.525	1.85	**0.26 ± 0.16**	0.18 ± 0.10	Hippurate	$1.5x10^{-9}$
2.685	1.81	0.16 ± 0.17	**0.25 ± 0.25**	Unidentified	$1.6x10^{-5}$
2.055	1.78	0.29 ± 0.04	**0.34 ± 0.06**	Unidentified	$6.0x10^{-23}$
1.285	1.74	0.11 ± 0.02	**0.15 ± 0.05**	Unidentified	$3.7x10^{-27}$
1.295	1.73	0.11 ± 0.02	**0.15 ± 0.05**	Unidentified	$6.4x10^{-25}$
7.425	1.70	**0.14 ± 0.05**	0.10 ± 0.04	N-phenylacetylglutamine	$2.1x10^{-25}$
2.035	1.70	0.31 ± 0.04	**0.35 ± 0.06**	N-acetylglutamate	$5.5x10^{-21}$
2.155	1.69	**0.37 ± 0.57**	0.25 ± 0.27	Unidentified	$3.4x10^{-2}$
7.125	1.68	**0.14 ± 0.19**	0.06 ± 0.09	Anserine	$6.0x10^{-7}$
7.265	1.66	0.21 ± 0.28	**0.15 ± 0.08**	Tryptophan	$6.6x10^{-13}$

[a)] Clusters were determined using two-step cluster analysis. Mean ± SD intensities of corresponding metabolites (all such values).

[b)] Peaks refer to the position of the metabolite in the nuclear magnetic resonance spectra.

[c)] Variable Importance in the Projection scores summaries the importance of the peaks in a Partial Least Square model.

[d)] P-value is based on independent t-tests. Multiple comparisons were corrected for using a Bonferroni adjustment. Bold values indicate significantly higher values. VIP, Variable Importance in the Projection. ppm, parts per million.

Table 3 Percentage energy contribution of food groups to total energy intake across the two clusters[a)]

Food groups[b)]	Cluster 1	Cluster 2	P[c)]
Rice, pasta, grains and flour	2.1 ± 3.1	2.1 ± 4.2	0.560
Savories	2.6 ± 5.2	3.4 ± 5.8	0.082
White bread and rolls	8.1 ± 6.9	8.0 ± 7.5	0.572
Wholemeal and brown bread	6.9 ± 6.2	6.0 ± 5.7	0.077
Breakfast cereals and porridge	**6.8 ± 5.5**	5.2 ± 5.0	$3.5\text{x}10^{-4}$
Biscuits, cakes and pastries	6.1 ± 6.2	5.9 ± 6.3	0.373
Wholemilk	2.6 ± 4.4	2.7 ± 3.9	0.572
Low fats and skimmed milks	**2.9 ± 3.5**	2.0 ± 3.1	0.002
Other milks/milk based drinks	0.5 ± 1.9	0.5 ± 2.1	0.709
Ice(creams),desserts	2.1 ± 3.2	1.9 ± 2.9	0.792
Cheeses	2.1 ± 2.7	2.5 ± 3.2	0.219
Yoghurts	1.7 ± 2.6	1.5 ± 2.4	0.127
Egg and egg dishes	1.5 ± 2.3	1.4 ± 2.1	0.361
Butter/fat spreads/cooking fats	3.5 ± 5.0	3.2 ± 4.0	0.997
Low fat spreads and oils	1.1 ± 2.0	1.0 ± 2.0	0.390
Potatoes	**3.4 ± 3.0**	2.8 ± 2.9	0.003
Chips and processed potatoes	3.3 ± 4.0	**5.1 ± 5.3**	$1.3\text{x}10^{-5}$
Vegetables/vegetable dishes	3.7 ± 3.0	3.6 ± 3.1	0.489
Fruit juices and smoothies	1.1 ± 1.7	1.0 ± 1.8	0.221
Fruit	**3.6 ± 3.6**	2.7 ± 3.3	$1.7\text{x}10^{-4}$
Savory snacks	1.8 ± 3.7	**2.4 ± 4.0**	0.036
Fish, fish dishes/products	**3.1 ± 3.9**	2.3 ± 3.4	0.003

Red meat	5.5 ± 5.2	5.9 ± 5.2	0.299
Poultry	2.4 ± 3.4	2.2 ± 3.1	0.230
Meat products	4.4 ± 5.5	**5.5 ± 6.0**	0.015
Red meat dishes	2.8 ± 4.4	2.8 ± 4.8	0.493
Poultry dishes	1.6 ± 3.6	1.7 ± 3.5	0.516
Alcoholic beverages	4.6 ± 6.0	6.1 ± 8.0	0.289
Sugars syrups preserves	1.8 ± 2.4	1.6 ± 2.5	0.244
Confectionary	2.4 ± 3.4	2.7 ± 3.9	0.335
Soups/sauces/condiments	2.5 ± 2.8	2.6 ± 2.8	0.462
Low energy beverages	0.1 ± 0.3	0.1 ± 0.3	0.931
High energy beverages	1.1 ± 2.4	**1.8 ± 3.1**	0.009

[a)] All values are Mean ± SD. Clusters were determined using two-step cluster analysis.

[b)] All food groups are based on percentage contribution.

[c)] P-values are based on Mann-Whitney U test. Bold values indicate significantly higher values.

Table 4 Comparison of mean daily nutrient intake between the clusters[a)]

	Cluster 1	Cluster 2	P[b)]
Carbohydrate % TE	**46.4 ± 7.1**	45.1 ± 6.8	0.033
Protein % TE	**17.7 ± 3.6**	16.7 ± 3.7	0.001
Sugars % TE	**18.8 ± 5.9**	17.6 ± 5.7	0.010
Starch % TE	26.8 ± 5.4	26.5 ± 5.1	0.638
Fat % TE	33.5 ± 6.2	34.2 ± 6.6	0.211
Saturated fat % TE	13.1 ± 3.6	13.3 ± 3.7	0.476
Monounsaturated Fat % TE	12.0 ± 2.7	**12.5 ± 2.6**	0.009
Polyunsaturated Fat % TE	6.1 ± 2.3	6.3 ± 2.6	0.120
Dietary fiber % TE	**4.2 ± 1.5**	3.8 ± 1.2	0.001
Alcohol % TE	4.0 ± 5.2	5.4 ± 7.1	0.188
Sodium (mg/10MJ)	2968.6 ± 649.0	3088.5 ± 787.9	0.265
Potassium (mg/10MJ)	**3917.9 ± 933.0**	3621.9 ± 775.1	$2.0x10^{-5}$
Calcium (mg/10MJ)	**1219.7 ± 501.7**	1058.4 ± 314.3	$2.0x10^{-4}$
Magnesium (mg/10MJ)	**375.6 ± 105.4**	340.6 ± 93.8	$1.7x10^{-6}$
Phosphorous (mg/10MJ)	**1739.2 ± 332.4**	1620.2 ± 314.3	$2.1x10^{-5}$
Iron (mg/10MJ)	**18.5 ± 18.4**	16.7 ± 18.9	0.022
Copper (mg/10MJ)	**1.8 ± 2.3**	1.5 ± 1.3	$1.3x10^{-5}$
Retinol (µg/10MJ)	**753.8 ± 1127.4**	564.9 ± 1617.7	$1.3x10^{-6}$
Carotene (µg/10MJ)	**5367.8 ± 4647.6**	4322.9 ± 3959.0	0.003
Vitamin D (µg/10MJ)	**7.5 ± 11.0**	5.7 ± 12.5	$1.7x10^{-6}$
Vitamin E (mg/10MJ)	**20.4 ± 52.9**	14.3 ± 34.3	0.003
Thiamine (mg/10MJ)	**4.7 ± 13.9**	3.5 ± 10.9	0.001
Niacin (mg/10MJ)	39.8 ± 56.3	33.4 ± 34.9	0.051

Vitamin B6 (mg/10MJ)	**6.0 ± 17.7**	5.2 ± 16.5	0.018
Vitamin B12 (µg/10MJ)	9.0 ± 11.7	**9.4 ± 26.1**	0.002
Folate (µg/10MJ)	**477.9 ± 288.4**	465.3 ± 654.7	$8.2x10^{-5}$
Riboflavin (mg/10MJ)	**4.7 ± 12.9**	3.7 ± 11.3	$1.4x10^{-4}$
Biotin (µg/10MJ)	**113.6 ± 860.6**	102.2 ± 823.4	$1.1x10^{-7}$
Vitamin C (mg/10MJ)	**188.0 ± 342.9**	152.2 ± 325.1	0.002

[a)] All values are Mean ± SD. All macronutrients are based on percentage energy contribution and micronutients are per 10MJ.

[b)]P-value are based on Mann-Whitney U test. Bold values indicate significantly higher values.

Table 5 Comparison of biochemical profiles across the two clusters[a)]

Biochemistry-core analytes	Cluster 1	Cluster 2	P Value[b)]
Serum Triglyceride (mmol/l)	1.29 ± 0.71	1.31 ± 0.71	0.746
Serum Total Cholesterol (mmol/l)	5.07 ± 1.03	5.01 ± 1.01	0.483
Serum Direct HDL (mmol/l)	**1.59 ± 0.42**	1.53 ± 0.46	0.031
Calculated LDL (mmol/l)	2.89 ± 0.91	2.88 ± 0.85	0.879
Serum Glucose (mmol/l)	5.35 ± 0.94	5.30 ± 0.87	0.503
Serum High Sentivity CRP (mg/l)	2.84 ± 2.86	2.88 ± 3.12	0.953
Serum Creatinine (umol/l)	90.83 ± 15.98	89.78 ± 13.97	0.549
Serum Potassium (mmol/l)	4.79 ± 0.44	4.77 ± 0.45	0.579
Serum Sodium (mmol/l)	144.11 ± 5.64	144.15 ± 6.13	0.964
Urinary Sodium (mmol/l)	91.11 ± 40.12	**102.39 ± 42.64**	0.002
Urinary Potassium (mmol/l)	45.60 ± 26.38	47.68 ± 28.46	0.470
Urinary Creatinine (umol/l)	9506.07 ± 5033.33	**12435.88 ± 6025.34**	1.7×10^{-9}
Serum Folate (nmol/l)	**36.90 ± 30.08**	29.32 ±20.34	3.3×10^{-5}
Red Cell Folate (nmol/l)	**1058.97 ± 453.72**	973.43 ± 452.97	0.009
Serum Vitamin B12 (pmol/l)	329.79 ± 228.03	327.51 ± 183.47	0.667
EGRac	1.34 ± 0.15	**1.38 ± 0.17**	0.004
Serum 25-hydroxyvitamin D (nmol/L)	63.00 ± 24.95	60.12 ± 24.34	0.112
PLP (nmol/L)	106.16 ± 96.03	100.79 ± 87.71	0.623
Plasma Homocysteine (mmol/l)	12.54 ± 3.34	12.46 ± 4.18	0.471
Haemoglobin (g/dl)	14.10 ± 1.34	14.32 ± 1.44	0.062
Serum NEFA (mmol/l)	0.69 ± 0.34	0.65 ± 0.35	0.109
Serum Insulin (µIU/ml)	8.90 ± 7.43	10.27 ± 9.30	0.053

[a] All values are Mean ± SD. Clusters were determined using two-step cluster analysis.

[b] P-values are based on independent t-tests (data was log transformed as required). Bold values indicate significantly higher values. EGRac, Erythrocyte glutathione reductase activation coefficient; PLP, Plasma Pyridoxal-5'-phosphate.

Supporting Information

Supporting Information Table 1 Description of standardized meals consumed by NutriTech food intake participants

Week 1		Week 2		Week 3	
Breakfast					
Female	**Male**	**Female**	**Male**	**Female**	**Male**
120g White bread	140g White bread	90g White bread	60g White bread	40g Brown bread	40g Brown bread
40g Eggs, boiled	60g Eggs, boiled	120 g Eggs, boiled	120 g Eggs, boiled	180g Eggs, boiled	200g Eggs, boiled
13 g Butter	20 g Butter	8g Butter	18 g Butter	4g Butter	12 g Butter
40 g Whole milk yogurt	125g Diet yogurt	110g Diet yogurt	330g Diet yogurt	310g Diet yogurt	600g Diet yogurt
250g Orange Juice	520g Orange Juice	220g Orange Juice	450g Orange Juice	50g Orange Juice	30g Orange Juice
Midday meal 1: Fish					
Female	**Male**	**Female**	**Male**	**Female**	**Male**
620g New potatoes	680g New potatoes	450g New potatoes	520g New Potatoes	270g New potatoes	300g New potatoes
28g Butter	31g Butter	32g Butter	37g Butter	37g Butter	42g Butter
50g Haddock	50g Haddock	100g Haddock	135g Haddock	210g Haddock	235g Haddock
100g Peas	120g Peas	150g Peas	125g Peas	170g Peas	180g Peas
Midday meal 2: Red Meat					
Female	**Male**	**Female**	**Male**	**Female**	**Male**
45g Braising steak	50g Braising steak	95g Braising steak	100g Braising steak	160g Braising steak	180g Braising steak
630g New potatoes	680g New potatoes	480g New potatoes	530g New potatoes	280g New potatoes	320g New potatoes
150g Carrots, boiled	250g Carrots, boiled	150g Carrots, boiled	250g Carrots, boiled	250g Carrots, boiled	250g Carrots, boiled
22g Butter	24 g Butter	22 g Butter	24 g Butter	20g Butter	22 g Butter
70g Gravy	100g Gravy	70g Gravy	100g Gravy	70g Gravy	100g Gravy
Midday meal 3: Processed Meat					
Female	**Male**	**Female**	**Male**	**Female**	**Male**
50g Ham	60g Ham	135g Ham	145g Ham	255g Ham	280g Ham
530g New potatoes	600g New potatoes	370g New potatoes	400g New potatoes	155g New potatoes	180g New potatoes
180g Baked beans	190g Baked beans	190g Baked beans	250g Baked beans	220g Baked beans	250g Baked beans
27g Butter	30 g Butter	30 g Butter	34 g Butter	30g Butter	35 g Butter
Midday meal 4: Chicken					
Female	**Male**	**Female**	**Male**	**Female**	**Male**
48g Chicken	52g Chicken	98g Chicken	110g Chicken	170g Chicken	190g Chicken
630g New potatoes	680g New potatoes	485g New potatoes	520g New potatoes	280g New potatoes	315g New potatoes
150g Carrots, boiled	250g Carrots, boiled	150g Carrots, boiled	250g Carrots, boiled	250g Carrots, boiled	250g Carrots, boiled
26g Butter	29g Butter	31g Butter	34g Butter	35g Butter	38g Butter
70g Gravy	100g Gravy	70g Gravy	100g Gravy	70g Gravy	100g Gravy
Midday meal 5: Vegetarian (Quorn)					
Female	**Male**	**Female**	**Male**	**Female**	**Male**
45g Braising steak	50g Braising steak	115g Quorn mince	125g Quorn mince	280g Quorn mince	305g Quorn mince
630g New potatoes	680g New potatoes	165g Baked potato	200g Baked potato	80g Baked potato	75g Baked potato
150g Carrots, boiled	250g Carrots, boiled	220g tomato based sauce	215g tomato based sauce	190g tomato based sauce	140g tomato based sauce
22g Butter	24 g Butter	40g Carrots	80g Carrots	65g carrots	75g Carrots
70g Gravy	100g Gravy	21g butter	21g butter	18g butter	22g Butter
		115g red peppers	115g red peppers	55g red peppers	95g red pepper
		20g cheddar cheese	30g cheddar cheese	31g cheddar cheese	32g Cheddar cheese
		75g Kidney beans	65g Kidney beans	35g red kidney beans	75g Kidney beans

7 pm meal 1: Fish					
Female	**Male**	**Female**	**Male**	**Female**	**Male**
510g New potatoes 22g Butter 40g Haddock 95g Peas	600g New potatoes 28g Butter 37g Haddock 140g Peas	345g New potatoes 26g Butter 80g Haddock 150g Peas	450g New potatoes 33g Butter 115g Haddock 140g Peas	200g New potatoes 30g Butter 160g Haddock 180g Peas	250g New potatoes 37g Butter 210g Haddock 180g Peas
7 pm meal 2: Red Meat					
Female	**Male**	**Female**	**Male**	**Female**	**Male**
36g Braising steak 500g New potatoes 150g Carrots, boiled 17g Butter 70g Gravy	45g Braising steak 630g New potatoes 150g Carrots, boiled 22g Butter 70g Gravy	75g Braising steak 380g New potatoes 150g Carrots, boiled 18 g Butter 70g Gravy	95g Braising steak 480g New potatoes 150g Carrots, boiled 22 g Butter 70g Gravy	128g Braising steak 210g New potatoes 250g Carrots, boiled 16g Butter 70g Gravy	160g Braising steak 280g New potatoes 250g Carrots, boiled 20g Butter 70g Gravy
7 pm meal 3: Processed Meat					
Female	**Male**	**Female**	**Male**	**Female**	**Male**
40g Ham 420g New potatoes 155g Baked beans 22g Butter	50g Ham 530g New potatoes 180g Baked beans 27g Butter	115g Ham 320g New potatoes 130g Baked beans 25 g Butter	135g Ham 370g New potatoes 190g Baked beans 30 g Butter	205g Ham 125g New potatoes 180g Baked beans 25g Butter	255g Ham 155g New potatoes 220g Baked beans 30g Butter
7 pm meal 4: Chicken					
Female	**Male**	**Female**	**Male**	**Female**	**Male**
38g Chicken 505g New potatoes 150g Carrots, boiled 21g Butter 70g Gravy	48g Chicken 630g New potatoes 150g Carrots, boiled 26g Butter 70g Gravy	78g Chicken 380g New potatoes 150g Carrots, boiled 25g Butter 70g Gravy	98g Chicken 485g New potatoes 150g Carrots, boiled 31g Butter 70g Gravy	138g Chicken 210g New potatoes 250g Carrots, boiled 28g Butter 70g Gravy	170g Chicken 280g New potatoes 250g Carrots, boiled 35g Butter 70g Gravy
7 pm Meal 5: Vegetarian (Quorn)					
Female	**Male**	**Female**	**Male**	**Female**	**Male**
36g Braising steak 500g New potatoes 150g Carrots, boiled 17g Butter 70g Gravy	45g Braising steak 630g New potatoes 150g Carrots, boiled 22g Butter 70g Gravy	130g Quorn mince 135g Baked potato 250g Tomato based sauce 170g carrots 19g Butter 60g red peppers 10g Cheese 18g Red Kidney beans	115g Quorn mince 170g Baked potato 220g tomato based sauce 40g Carrots 20g butter 115g red peppers 22g cheddar cheese 75g Kidney beans	250g Quorn mince 60g Baked potato 180g tomato based sauce 50g carrots 17g butter 40g red peppers 20g cheddar cheese 30g red kidney beans	290g Quorn mince 80g Baked potato 190g tomato based sauce 65g Carrots 19g Butter 55g red pepper 31g Cheddar cheese 35g Kidney beans

Supporting Information Table 2 NutriTech participant characteristics[a)]

Characteristic	NutriTech population (n=50)
Gender	25 M 25 F
BMI (kg/m^2)	28.6 ± 3.5
Age (y)	60 ± 4
Systolic BP (mm/Hg)	132.4 ± 15.5
Diastolic BP (mm/Hg)	77.9 ± 10.7

a)Mean ± SD (all such values). BP, Blood Pressure.

Supporting Information Table 3 Foods and beverages absent from the NutriTech participant's (n=49) controlled diet during the food intake study[a)]

Food groups	Food Intake Study
Rice pasta grains flour starch	N[b)]
Savories	N
Wholemeal & brown bread rolls	N
Breakfast cereals & porridge	N
Biscuits, cakes & pastries	N
Wholemilk	N
Low fats and skimmed milks	N
Other milks, milk based beverages & other beverages	N
Creams, ice creams and desserts	N
Low fat spreads and oils	N
Chips and processed potatoes	N
Fruit	N
Savory snacks	N
Red meat dishes	N
Alcoholic beverages	N
Sugars syrups preserves & sweeteners	N
Confectionary	N
Low energy beverages	N
High energy beverages	N

[a)] Food intake study represents food intake during week 2 of the controlled NutriTech study.

[b)] N, no this food group was not consumed.

Supporting Information Table 4 Prediction of cluster membership using a test set[a)]

Cluster number	Number of subjects who were correctly placed		Total	% correctly placed[b)]
Cluster 1	91		99	91.9
Cluster 2		90	90	90.0

[a)] The cluster membership of each sample in the NANS test set (n=189) was predicted using the NANS training set model.

[b)] Represents the percentage of subjects correctly classified into their corresponding cluster.