Potential of food intake biomarkers in nutrition research

Aoife E. McNamara [1, 2], and Lorraine Brennan [1, 2]

[1] UCD School of Agriculture and Food Science, Institute of Food and Health, UCD, Belfield, Dublin 4, Ireland

[2] UCD Conway Institute, UCD, Belfield, Dublin 4, Ireland

Corresponding Author:

Professor Lorraine Brennan

UCD School of Agriculture and Food Science,

UCD Institute of Food and Health,

UCD, Belfield, Dublin 4, Ireland.

Email: lorraine.brennan@ucd.ie

Phone: 00 353 1 7166815

Short Title: Potential of food intake biomarkers

Keywords: biomarkers, dietary assessment, food intake, metabolomics

Abstract

The influence of dietary habits on health/ disease is well-established. Accurate dietary assessment is essential to understand metabolic pathways/ processes involved in this relationship. In recent years, biomarker discovery has become a major area of interest for improving dietary assessment. Well-established nutrient intake biomarkers exist, however, there is growing interest in identifying and using biomarkers for more accurate and objective measurements of food intake.

Metabolomics has emerged as a key tool used for biomarker discovery, employing techniques such as nuclear magnetic resonance spectroscopy, or mass-spectrometry. To date, a number of putatively identified biomarkers were discovered for foods including meat, cruciferous vegetables and legumes. However, many of the results are associations only and lack the desired validation including dose-response studies.

Food intake biomarkers can be employed to classify individuals into consumers/ non-consumers of specific foods, or into dietary patterns. Food intake biomarkers can also play a role in correcting self-reported measurement error, thus improving dietary intake estimates. Quantification of food intake was previously performed for citrus (proline betaine), chicken (guanidoacetate) and grape (tartaric acid) intake. However, this area still requires more investigation and expansion to a range of foods.

The present review will assess current literature of identified specific food intake biomarkers, their validation and the variety of biomarker uses. Addressing utility of biomarkers and highlighting gaps in this area is important to advance the field in the context of nutrition research.

Introduction

It is well established that environmental and lifestyle factors, such as dietary intake and habits, influence health and disease outcomes [1]. Epidemiological evidence has reported associations between dietary intake and positive health effects for cardiovascular disease (CVD) [2, 3], diabetes [4], and certain cancers [5-7]. In order to interpret the effect of diet on health it is critical to accurately measure an individual's, or a population's, dietary intake. Traditional self-reported dietary assessment techniques, including Food Frequency Questionnaires (FFQs), Dietary Recalls and Weighed Food Records, are subject to well documented limitations. For example, self-reported methods are at risk of reporting inaccuracy, subjective estimation of portion sizes, recall bias and misreporting [8-11]. Consequently, there is a need for the development of more accurate and objective dietary assessment measures, such as dietary biomarkers.

The discovery of dietary biomarkers is an area of increasing interest. Presently, there are only a few biomarkers for dietary assessment that are well-established, capturing intake of salt, protein, sucrose and fructose [1]. Twenty-four-hour urinary nitrogen is a well-known biomarker of protein [1] and is often used to validate self-reported intake [12], or to compare the accuracy of two dietary assessment methods [13]. Urinary concentrations of sucrose and fructose are dose-responsive and predictive biomarkers of dietary sugars [14,15] However, many studies investigating sucrose/ fructose as a potential intake biomarker have been observational, with associations appearing low to moderate [16], perhaps due to between/ within subject variation in urinary sucrose/ fructose absorption, tissue uptake and excretion [14]. While these biomarkers are accepted as more accurate and useful, they are reflective of dietary nutrient habits instead of consumption of specific foods, highlighting the need for food intake biomarkers.

A newly defined flexible classification scheme for biomarkers related to food intake was recently published [17]. The authors outline six subclasses of dietary and health biomarkers to be included under the previously suggested major classes of biomarker: Exposure, Effect and Susceptibility. Four of these subclasses are associated with dietary intake and are as follows: 1) Food compound intake biomarkers: nutrients or non-nutrients reflective of dietary intake, 2) Food intake biomarkers: nutrients or non-nutrients reflective of intake of a specific food, 3) Dietary pattern biomarkers: A set of food intake biomarkers that can distinguish between different dietary habits or indicate a high adherence to a pre-defined diet (e.g. Mediterranean or Nordic diets), and 4) Food compound status biomarkers: nutrients and non-nutrients indicating accumulated stores of compounds in the body. The final two subclasses (effect and physiological markers) are not products of dietary intake and therefore are not covered by this review. Biomarkers of food intake can be single metabolites, or a

combination of metabolites, reflecting the consumption of either a specific food or food group, displaying a clear time- and dose-response after intake [17].

Metabolomic techniques for food intake biomarker discovery

Through use of metabolomics a number of food intake biomarkers have emerged in the literature. At the most broadest definition, metabolomics is the study of endogenous or exogenous metabolites in a biological sample. The human metabolome is influenced by multiple factors such as genetics, the microbiome [18], and environmental factors including diet and lifestyle [19]. Analysis of metabolites is usually performed using nuclear magnetic resonance (NMR) spectroscopy, or mass-spectrometry (MS), which can be coupled with a separation technique such as liquid (LC) or gas chromatography (GC) [20, 21].

In brief, NMR is a popular metabolomics platform frequently employed for the discovery and identification of novel food intake biomarkers. NMR captures quantitative metabolite data in a robust fashion. It is a non-destructive method, relatively fast and requires little sample preparation [22]. NMR has a comparatively lower sensitivity and requires larger sample volumes compared to other analytical techniques like MS methods of analysis [23]. However, the reproducibility across multiple laboratories of NMR analysis is very high. This technique is useful for broad-based analyses and high abundance metabolites [24]. Mass-spectrometry based techniques are extremely sensitive and can analyse small sample volumes, however samples are non-recoverable after analysis. Furthermore, sample preparation is more laborious than NMR. Mass-spectrometry techniques coupled to LC or GC separate compounds based on their physiochemical properties, which are eluted at various retention times (RT). These compounds are then ionised, determining their mass-to-charge ratio (m/z) [25]. Compound identification can be made by combining RT information with m/z along with additional analyses, such as fragmentation patterns from tandem MS, to compare against standards and spectral libraries. Different chromatographic techniques, coupled to MS, can identify different metabolites. Examples of routinely measured metabolites are polar and volatile compounds, amino acids, biogenic amines, peptides, intact lipids organic acids, bile acids, and fatty acids as well as other macromolecules [26, 27].

In the context of applying metabolomics to food intake studies a number of challenges exist. These include the generation of a large amount of data to be processed and identification of metabolites to a high confidence level [22]. Currently, identification of metabolites is reliant on the availability of analytical standards for confirmation and the availability of comprehensive spectral libraries and

databases. Unfortunately, many such databases contain few food related compounds. Notwithstanding these challenges, biomarkers have been identified for a number of foods.

Study designs to identify food intake biomarkers

To date, studies were performed to identify potential food intake biomarkers for multiple foods and food groups, covering a wide range of components of the human diet [28-32]. Numerous study designs can be employed to identify food intake biomarkers. Previous successful designs include acute intervention studies, short/medium term interventions and cross-sectional cohort studies [28, 33]. The intervention study designs involve consumption of specific food(s) over a defined period of time and biofluids, such as blood and urine, are collected at specific time-points depending on research interests. Human intervention studies make it possible to control potential confounding factors and allow the focused investigation of the effect of specific food/ food group intake on biological samples. However, intervention studies are often performed in smaller sample sizes, and results may not be directly applicable to free-living populations. In order to overcome this limitation, it is important that identified potential food intake biomarkers are validated in other, larger, less-controlled populations.

Using samples from epidemiology studies, enables examination of the relationships between self-reported food intake and biomarkers measured in urine or blood samples. Epidemiological studies collect dietary data from large sample sizes, are relatively low burden on participants and more likely to be indicative of a free-living setting. However, because of the uncontrolled setting there is the potential for confounding variables. One of the potential limitations of using epidemiological data for food intake biomarker identification is that biomarkers identified may be present in more than one food or food group, further highlighting the need for biomarker validation. Ultimately, the design of the research will be guided by the research question, taking into consideration the limitations of each approach. The choice of biofluid examined is also important as some metabolites will appear exclusively, or are more concentrated, in some biofluids [26]. There are multiple biofluids which can be used for food intake biomarkers identification, however urine and blood samples are most frequently employed as they are easily accessible and contain numerous compounds of biological importance, including food intake metabolites [22, 34].

Validation of food intake biomarkers

Currently, there are extensive research efforts in the identification of food intake biomarkers, however, efforts in validation of the biomarkers are still lacking. To address this issue a number of criteria were recently developed for the validation of food intake biomarkers (Figure 1). The criteria include the following eight points: plausibility, dose-response, time-response, robustness, reliability,

stability, analytical performance and reproducibility [35]. Examining the plausibility of a food intake biomarker includes confirming food specificity and establishing any food chemistry/ food processing/ experimental explanations for increased concentration after consuming the food. The food intake biomarker's response to different portions of specific food should be examined, taking into account a range of intakes, habitual baseline levels, bioavailability, excretion timeline and saturation levels. The biomarker's time-response, half-life and kinetics of the biomarker are explored both after a single exposure and repeated measures over time, examining its stability as an estimate of longer-term intake. Biomarkers must be robust by demonstrating suitability in multiple free-living populations, and any food interactions identified. Investigating the reliability of a biomarker requires comparing the biomarker with a gold standard, other biomarkers of the food or other dietary assessment methods which provide good measure of true exposure. To be effective in nutrition research food intake biomarkers chosen must be stable within the biofluid used for analysis. The analytical performance of a biomarker must be well-documented, its precision, accuracy and detection limits, and any inter- / intra-batch variation assessed. The results of a biomarker's performance and efficacy should be reproducible with validated methods established for comparing results across different laboratories. Applying these validation criteria to the large number of potential food intake biomarkers will allow for the development of robust and valid biomarkers.

Applications of food intake biomarkers in nutrition research

At present, there are a number of putative specific food intake biomarkers identified, with varying levels of fulfilled validation criteria. Furthermore, there is lack of research which demonstrates the multitude of applications of these biomarkers. Food intake biomarkers are extremely useful tools which can not only determine dietary exposure but be applied to correct for self-reported measurement error and the classification of dietary patterns.

Using food intake biomarkers to classify intake

There are many examples in the literature where biomarkers were used to classify individuals into consumers or non-consumers of specific foods (Table 1). In the INTERMAP study urinary proline betaine was used to classify participants as citrus consumers or non-consumers [30]. Using receiver operating characteristics (ROC) curves proline betaine was able to identify citrus consumers with a specificity and sensitivity of 92.3% and 80.6% respectively. A study investigating Nordic diets was able to differentiate between consumers and controls of specific plant foods (cabbage, beetroot, strawberries and walnuts) based on peak areas of potential food intake biomarkers identified in 24 hour urine by UPLC-qTOF-MS [36]. Urine samples from the SU.VI.MAX2 study were used to identify coffee intake biomarkers, many of the identified biomarkers performed well at separating samples

from high and low coffee consumers [32]. Atractyligenin glucuronide had the highest ROC area under the curve (AUC) and outperformed caffeine (AUC = 0.95 vs 0.72 respectively). As part of the WHOLEheart study, alkylresourcinols (ARs), biomarkers of wholegrain intake, were quantified in plasma samples by GC-MS. Plasma concentrations were significantly different between the control (low wholegrain intake ,<30g/d) and intervention groups (high wholegrain intake, 60g/d or 120g/d) ($p \leq 0.0073$ across analyses) demonstrating they could distinguish between consumers and non-consumers [37]. These plasma ARs concentrations were also capable of distinguishing quartile of wholegrain intake at a slight to fair level (misclassification rate of 9% - 12%) [37]. A recently published paper examined non-fasting serum samples to identify the most predictive biomarkers for 42 food items or food groups using ROC AUCs to separate high and low consumers by quintiles of intake [38]. The average AUC was 0.75 (ranging from 0.65-0.98), however, the authors were unable to distinguish metabolites which were food intake biomarkers and metabolites resulting from diet-induced changes in metabolism.

Food intake biomarkers can also be combined to achieve or improve classification of dietary intake. A recently published study used five discriminative metabolites to classify high and non-consumers of banana [39]. The predictive ability of these metabolites was tested using ROC curve analysis using Partial Least Squares- Discriminant Analysis models of biomarker combinations. The combination of all five metabolites was highly predictive (AUC = 0.90; error rate (ER) = 0.13) for high banana consumers versus non-consumers, however it was a combination of just two of these metabolites which performed the best overall at classifying recent banana high consumers (AUC = 0.92; ER_{test} = 0.11). Work from our own research lab identified four food intake biomarkers of sugar sweetened beverages using heat-map analysis of metabolomic urinary profiles from the National Adult Nutrition Survey (NANS) study [40]. These markers were combined in a panel and ROC curves demonstrated that the panel could discriminate between consumers and non-consumers of sugar-sweetened beverages (AUC = 0.8) and was more predictive of intake than the individual biomarkers themselves (AUCs ranging from 0.5-0.7). A multimetabolite biomarker panel, made up of beer ingredient and food processing biomarkers, was capable of distinguishing beer consumption from urine samples collected before and up to 12 hours after intake of beer with excellent efficiency (AUC = 1) [41]. Using two food intake biomarkers of wine, analysis of PREDIMED study data revealed a stepwise logistic regression model capable of identifying wine consumers compared to non-consumers (AUC = 0.92) and detecting these consumers up to three days after the last glass of wine [42]. Fasting urine metabolomic data from the PREDIMED study analysed by LC-MS was also used to develop a multimetabolite panel capable of predicting non-bread consumers and whole-grain bread consumers

(ROC AUCs > 0.93 for both positive and negative mode models) [43]. The multimetabolite panel contained ARs, benzoxazinoids, microbial metabolites, exogenous metabolites and a heat-treatment product. The same authors also developed a panel of urinary food intake biomarkers for discriminating cocoa consumers from non-consumers in the same population (ROC AUC = 0.93) [44].

The above-mentioned studies demonstrate that food intake biomarkers can be very efficient at classifying consumers and non-consumers of specific foods and they have the potential to be used to validate self-reported findings. However, this approach is qualitative and further research into these biomarkers is necessary to enable the field to move from qualitative to quantitative assessment of food intakes.

Quantifying intake using food intake biomarkers

Examining a biomarker's ability to quantify intake can progress food intake biomarkers beyond the dichotomous classification of consumers and non-consumers. Previous work from our research group examined the potential of the well-established marker of citrus intake, proline betaine, in determining citrus intake [45]. Employing calibration curves developed from a controlled dietary intervention study (NutriTech), urinary proline betaine concentrations were used to determine the citrus intake in an independent cross-sectional study of 565 individuals. There was excellent agreement between the self-reported intake (estimated from a four day semi-weighed dietary record) and the biomarker-estimated intake with a low mean bias of 4.3 g between methods. This study clearly demonstrates the potential of well validated food intake biomarkers. Our research group also applied a similar approach to a biomarker of chicken intake– guanidoacetate [46]. Urinary guanidoacetate demonstrated a dose-response relationship with increasing chicken intake in the NutriTech study and a calibration curve developed was able to discriminate between high and non-consumers of chicken in an independent cross-sectional study. Guanidoacetate demonstrated good agreement between self-reported and biomarker-estimated intake with a low mean bias of -30.2 g between methods. Garcia-Perez and colleagues established a dose-response relationship between grape intake and urinary tartaric acid levels and investigated the ability of tartaric acid to determine grape intake [47]. The agreement between estimated intake and actual intake was good and a correlation coefficient of $R^2 = 0.9$ was reported. Overall, these three examples, summarised in Table 2, provide strong evidence of the potential of food intake biomarkers to quantify intake of specific foods and demonstrate the importance of assessing dose-response relationships of identified biomarkers. While these examples support the potential of biomarkers for quantification of food intake there are a number of limitations worth mentioning. The above studies are reliant on well-controlled feeding studies to estimate the

relationships between biomarkers and intake. Not all biomarkers, will exhibit a linear relationship with intake thus limiting their potential to predict intake.

Developing calibration equations to correct dietary data

Another application of food intake biomarkers is the development of calibration equations that can correct self-reported intake data. Previously, this approach has been applied to nutrient data from the Women's Health Initiative and was used to develop biomarker-calibrated equations which uncovered disease associations that were not identified in uncalibrated data [48]. Work from our research group implemented a similar approach using food intake biomarker data to develop calibration equations utilising self-reported intakes and biomarker-derived estimates of citrus intakes from the NANS [49]. Statistical transformations were performed on the data to achieve optimal calibration specifications, which were then applied to correct for the error in self-reported intake data and achieve a more accurate and objective measure of true intake. This work is very promising, demonstrating the utility of food intake biomarkers in nutrition research, however further investigation is required. Application of this method to other food intake biomarkers would enable the correction of self-reported data in large epidemiological studies and improve dietary assessment. This research also developed a framework for determining the amount of biomarker data that would be required to correct for self-reported error in epidemiological studies, as it is not always feasible to collect biofluids from all subjects [49]. Results indicated that biomarker data from approximately 20-30 % of subjects would be sufficient to correct for errors. This important finding will allow improvement in accuracy of dietary intake estimation, especially in larger study sample sizes.

Biomarker-based classification of dietary patterns

Analysis of dietary patterns allows researchers to gain a broader insight into dietary intake and habits as opposed to a focus on specific foods. Dietary pattern analysis encompasses the quantities, proportion, variety and combination of foods/ beverages consumed as well as the frequency of consumption [50]. The ability of metabolomics and food intake biomarkers to classify dietary patterns or monitor adherence to pre-defined diets has been investigated using a range of different study designs (Table 3). Application of interventions studies has elegantly demonstrated that metabolomic profiles can distinguish different dietary patterns. Untargeted metabolomic profiles were employed to distinguish between two Nordic dietary patterns used in an intervention study; the New Nordic Diet (NND) or an Average Danish Diet (ADD) [51]. A multivariate model was established using urinary metabolome profiles, which classified the two dietary patterns with a low misclassification error rate (19 %). A follow up paper, using a classification model built on plasma metabolic profiles, was capable of assessing dietary pattern compliance between NND and ADD (Average ROC AUCs

for positive and negative mode = 0.88 and 0.74 respectively) [52]. Similarly, a plasma metabolome-based dietary pattern classification model was performed by Esko and colleagues on date from a feeding study [53]. Three different diets with varying macronutrient compositions (Low fat- 60 % carbohydrate, 20 % fat, 40 % protein, Low glycaemic index- 40 % carbohydrate, 40 % fat, 20 % protein and Very low carbohydrate- 10 % carbohydrate, 60 % fat and 30 % protein) were distinguishable using this model. The models were able to identify which dietary pattern participants were following in 95% of cases in the test set [53]. In a separate intervention study a fasting serum metabolite panel was identified, that could distinguish between participants consuming a "Dietary Approaches to Stop Hypertension (DASH)" diet, a fruit and vegetable diet or a control diet. Predictability of the model was examined in a test set reporting a C statistic (AUC) of 0.961 indicating good ability to classify individuals into the dietary pattern followed [54]. Using a controlled intervention Garcia-Perez and colleagues developed a model based on urinary metabolomics data that could classify individuals into dietary patterns. The four diets were based on the World Health Organisation (WHO) healthy eating guidelines for the prevention of non-communicable diseases (NCDs) [55]. The model was validated in two separate population groups. Collectively, these studies provide strong evidence for biomarker-based metabolomic profiling to classify and monitor adherence to dietary patterns.

Further evidence is also available from studies performed using cross-sectional data. A recent study demonstrated the ability of ^{1}H NMR profiles to distinguish dietary habits related to varying degrees of meat consumption/ avoidance [56]. Serum metabolite profiles were capable of correctly classifying 97.5% of meat eaters compared to non- meat eaters (ROC AUC = 1) and, inversely, 92.5% of vegans compared to non-vegans (ROC AUC = 0.98). Work has emerged to support the potential measurement of adherence to pre-defined dietary patterns such as the Mediterranean Diet through metabolomic profiles [57, 58, 59]. Macias and colleagues identified fasting plasma metabolites capable of discriminating between low and high Mediterranean Diet Score (MDS) and their correlations with food intakes (ROC AUC = 0.74) [57]. In a study of postmenopausal women in the US, metabolite levels in serum samples were capable of predicting low and high adherence to four healthy diet scores (the alternate Mediterranean diet score (aMED), alternate Healthy Eating Index (AHEI)-2010, DASH diet, and the Healthy Eating Index (HEI)-2015) [58]. Examining a test dataset, revealed that the serum metabolites discriminated between high and low quintiles of adherence to the four different healthy dietary pattern scores (ROC AUCs $\geq$ 0.76 for each diet score individually) [58]. Additionally, a fasting plasma metabolite score was correlated with adherence to Mediterranean Diet (Spearman's p = 0.42)

in a UK population [59]. Overall, these studies add to the evidence base supporting the relationship between metabolites and dietary patterns.

Finally, evidence has also emerged to support the ability of a biomarkers-based approach to determine de novo dietary patterns and classify individuals into these. Research from our group, identified two distinct dietary patterns (a healthy and unhealthy pattern) using only urinary metabolomic profiles (n=567). Using this model in an independent study revealed that 94% of subjects were correctly classified into the correct dietary pattern group [60]. A recent study used reduced rank regression to identify dietary patterns reflecting metabolites that were pre-selected to be associated with a disease [61]. The approach identified three dietary patterns and represents an interesting approach for examining disease-related metabolites and dietary patterns.

Collectively, the emerging data from both intervention and cross sectional studies support the concept that metabolomic-based food intake biomarker models could be used to identify dietary patterns and to further examine the relationship between dietary patterns and health outcomes in larger epidemiological studies.

Future outlook

While the majority of work to date has focused on the identification of new biomarkers of intake there are promising examples of how these biomarkers could be used in nutrition research. The current limitations in applications of food intake biomarkers stem from the limited number of robust biomarkers already demonstrated in the literature, therefore future work in this area needs to focus on identifying specific and sensitive food intake biomarkers and validating them according to recently outlined criteria [35]. Food intake biomarkers have the potential to improve the accuracy of dietary assessment methods. Biomarkers can be used for the correction of measurement error in dietary data collection using statistical transformations. Metabolomic profiles can be used to classify adherence to dietary patterns and habits in large sample sizes and monitor compliance to study protocols or medically prescribed diets. This improved dietary information can help to unravel the impact and interactions of dietary components in metabolic processes and pathways, as well as elucidating the relationship between diet and disease outcome.

328 **Acknowledgements**

329 The authors would like to thank the Irish section of the Nutrition Society for inviting the present
330 review paper as part of the postgraduate review competition.

331 **Financial support**

332 This work was supported by a H2020 European Research Council (647783).

333 **Conflict of interest**

334 None.

335 **Figure Legends**

336 **Figure 1.** Outline of the recently developed criteria for the validation of food intake biomarkers
337 (adapted from Dragsted *et al.,* (2018) [35]).

338

References

1. Bingham SA. (2002) Biomarkers in nutritional epidemiology. *Public Health Nutr* **5**, 821-7.
2. Nuñez-Cordoba JM, Alonso A, Beunza JJ, *et al.* (2009) Role of vegetables and fruits in Mediterranean diets to prevent hypertension. *Eur J Clin Nutr.* **63**, 605-12.
3. Panagiotakos DB, Pitsavos C, Kokkinos P, *et al.* (2003) Consumption of fruits and vegetables in relation to the risk of developing acute coronary syndromes; the CARDIO2000 case-control study. *Nutr J.* **2,** doi:10.1186/1475-2891-2-2.
4. Nöthlings U, Schulze MB, Weikert C, *et al.* (2008) Intake of vegetables, legumes, and fruit, and risk for all-cause, cardiovascular, and cancer mortality in a European diabetic population. *J Nutr.* **138,** 775-81.
5. World Health Organization (2003) *Diet, Nutrition and the Prevention of Chronic Diseases.* Joint WHO/FAO Expert Consultation. WHO Technical Report Series no. 916. Geneva: WHO.
6. Vieira AR, Vingeliene S, Chan DS, *et al.* (2015) Fruits, vegetables, and bladder cancer risk: a systematic review and meta-analysis. *Cancer Med.* **4,** 136-46.
7. Gullett NP, Mazurak VC, Hebbar G, *et al.* (2011) Nutritional interventions for cancer-induced cachexia. *Curr Probl Cancer.* **35,** 58-90.
8. Arija V, Abellana R, Ribot B, *et al.* (2015) Biases and adjustments in nutritional assessments from dietary questionnaires. *Nutr Hosp.* **31,** 113-8.
9. Dhurandhar NV, Schoeller D, Brown AW, *et al.* (2015) Energy balance measurement: when something is not better than nothing. *Int J Obes (Lond).* **39,** 1109-13.
10. Kipnis V, Midthune D, Freedman L, *et al.* (2002) Bias in dietary-report instruments and its implications for nutritional epidemiology. *Public Health Nutr.* **5,** 915-23.
11. Shim JS, Oh K, Kim HC. (2014) Dietary assessment methods in epidemiologic studies. *Epidemiol Health.* **36,** doi: 10.4178/epih/e2014009.
12. Rasmussen LG, Winning H, Savorani F, *et al.* (2012) Assessment of the effect of high or low protein diet on the human urine metabolome as measured by NMR. *Nutrients.* **4,** 112-31.
13. McKeown NM, Day NE, Welch AA, *et al.* (2001) Use of biological markers to validate self-reported dietary intake in a random sample of the European Prospective Investigation into Cancer United Kingdom Norfolk cohort. *Am J Clin Nutr.* **74,** 188-96.
14. Tasevska N. (2015) Urinary Sugars--A Biomarker of Total Sugars Intake. *Nutrients.* **7**, 5816-33.

15. Luceri C, Caderni G, Lodovici M, *et al.* (1996) Urinary excretion of sucrose and fructose as a predictor of sucrose intake in dietary intervention studies. *Cancer Epidemiol Biomarkers Prev*. **5,** 167-71.
16. Davy B, Jahren H. (2016) New markers of dietary added sugar intake. *Curr Opin Clin Nutr Metab Care.* **19,** 282-8.
17. Gao Q, Praticò G, Scalbert A, *et al.* (2017) A scheme for a flexible classification of dietary and health biomarkers. *Genes Nutr.* **12**, 1-15.
18. Fujisaka S, Avila-Pacheco J, Soto M, *et al.* (2018) Diet, Genetics, and the Gut Microbiome Drive Dynamic Changes in Plasma Metabolites. *Cell Rep.* **22,** 3072-86.
19. Zamora-Ros R, Achaintre D, Rothwell JA, *et al.* (2016) Urinary excretions of 34 dietary polyphenols and their associations with lifestyle factors in the EPIC cohort study. *Sci Rep.* **6**, doi:10.1038/srep26905.
20. Scalbert A, Brennan L, Manach C, *et al.* (2014) The food metabolome: a window over dietary exposure. *Am J Clin Nutr.* **99,** 1286-308.
21. Gibbons H, O'Gorman A, Brennan L. (2015) Metabolomics as a tool in nutritional research. *Curr Opin Lipidol.* **26**, 30-4.
22. Ulaszewska MM, Weinert CH, Trimigno A, *et al.* (2019) Nutrimetabolomics: An Integrative Action for Metabolomic Analyses in Human Nutritional Studies. *Mol Nutr Food Res.* **63,** doi: 10.3390/metabo9110261.
23. Wishart DS. (2008) Metabolomics: applications to food science and nutrition research. *Trends Food Sci Technol.* **19**, 482-93.
24. Pan Z, Raftery D. (2007) Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. *Anal Bioanal Chem.* **387,** 525-7.
25. Sparkman OD, Penton Z, Kitson F. (2011) Chapter 4: Mass Spectrometry Instrumentation, In *Gas Chromatography and Mass Spectrometry: A Practical Guide.* 2nd ed. pp. 89-148, Oxford, UK: Academic Press Elsevier
26. Bouatra S, Aziat F, Mandal R, *et al.* (2013) The human urine metabolome. *PLoS One.* **8,** doi.org/10.1371/journal.pone.0073076.
27. Griffin JL, Atherton H, Shockcor J, *et al.* (2011) Metabolomics as a tool for cardiac research. *Nat Rev Cardiol.* **8,** 630-43.
28. Cross AJ, Major JM, Sinha R. (2011) Urinary biomarkers of meat consumption. *Cancer Epidemiol Biomarkers Prev*. **20,** 1107-11.
29. Lloyd AJ, Favé G, Beckmann M, *et al.* (2011) Use of mass spectrometry fingerprinting to identify urinary metabolites after consumption of specific foods. *Am J Clin Nutr*. **94**, 981-91.

30. Heinzmann SS, Brown IJ, Chan Q, *et al.* (2010) Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. *Am J Clin Nutr*. **92,** 436-43.
31. Edmands WM, Beckonert OP, Stella C, *et al.* (2011) Identification of human urinary biomarkers of cruciferous vegetable consumption by metabonomic profiling. *J Proteome Res.* **10,** 4513-21.
32. Rothwell JA, Fillâtre Y, Martin JF, *et al.* (2014) New biomarkers of coffee consumption identified by the non-targeted metabolomic profiling of cohort study subjects. *PLoS One.* **9,** doi.org/10.1371/journal.pone.0093474.
33. O'Gorman A, Gibbons H, Brennan L. (013) Metabolomics in the identification of biomarkers of dietary intake. *Comput Struct Biotechnol J.* **4,** doi: 10.5936/csbj.201301004.
34. Guasch-Ferré M, Bhupathiraju SN, Hu FB. (2018) Use of Metabolomics in Improving Assessment of Dietary Intake. *Clin Chem*. **64,** 82-98.
35. Dragsted LO, Gao Q, Scalbert A, *et al.* (2018) Validation of biomarkers of food intake-critical assessment of candidate biomarkers. *Genes Nutr*. **13,** doi: 10.1186/s12263-018-0603-9.
36. Andersen MB, Kristensen M, Manach C, *et al.* (2014) Discovery and validation of urinary exposure markers for different plant foods by untargeted metabolomics. *Anal Bioanal Chem.* **406,** 1829-44.
37. Ross AB, Bourgeois A, Macharia HN, *et al.* (2012) Plasma alkylresorcinols as a biomarker of whole-grain food consumption in a large population: results from the WHOLEheart Intervention Study. *Am J Clin Nutr.* **95**, 204-11.
38. Wang Y, Gapstur SM, Carter BD, *et al.* (2018) Untargeted Metabolomics Identifies Novel Potential Biomarkers of Habitual Food Intake in a Cross-Sectional Study of Postmenopausal Women. *J Nutr.* **148,** 932-43.
39. Vázquez-Manjarrez N, Weinert CH, Ulaszewska MM, *et al.* (2019) Discovery and Validation of Banana Intake Biomarkers Using Untargeted Metabolomics in Human Intervention and Cross-sectional Studies. *J Nutr.* **149,** 1701-1713
40. Gibbons H, McNulty BA, Nugent AP, *et al.* (2015) A metabolomics approach to the identification of biomarkers of sugar-sweetened beverage intake. *Am J Clin Nutr.* **101,** 471-7.
41. Gürdeniz G, Jensen MG, Meier S, *et al.* (2016) Detecting Beer Intake by Unique Metabolite Patterns. *J Prot Res.* **15,** 4544-56.
42. Vázquez-Fresno R, Llorach R, Urpi-Sarda M, *et al.* (2015) An NMR metabolomics approach reveals a combined-biomarkers model in a wine interventional trial with validation in free-living individuals of the PREDIMED study. *Metabolomics.* **11,** 797-806.

438 43. Garcia-Aloy M, Llorach R, Urpi-Sarda M, *et al.* (2015) Nutrimetabolomics fingerprinting to
439 identify biomarkers of bread exposure in a free-living population from the PREDIMED study
440 cohort. *Metabolomics.* **11,** 155-65.
441 44. Garcia-Aloy M, Llorach R, Urpi-Sarda M, *et al.* (2015) A metabolomics-driven approach to
442 predict cocoa product consumption by designing a multimetabolite biomarker model in free-
443 living subjects from the PREDIMED study. *Mol Nutr Food Res*. **59,** 212-20.
444 45. Gibbons H, Michielsen CJR, Rundle M, *et al.* (2017) Demonstration of the utility of biomarkers
445 for dietary intake assessment; proline betaine as an example. *Mol Nutr Food Res.* **61,**
446 doi.org/10.1002/mnfr.201700037.
447 46. Yin X, Gibbons H, Rundle M, *et al.* (2017) Estimation of Chicken Intake by Adults Using
448 Metabolomics-Derived Markers. *J Nutr.* **147,** 1850-1857
449 47. Garcia-Perez I, Posma JM, Chambers ES, *et al.* (2016) An Analytical Pipeline for Quantitative
450 Characterization of Dietary Intake: Application To Assess Grape Intake. *J Agric Food Chem*.
451 **64,** 2423-31.
452 48. Tinker LF, Sarto GE, Howard BV, *et al.* (2011) Biomarker-calibrated dietary energy and
453 protein intake associations with diabetes risk among postmenopausal women from the Women's
454 Health Initiative. *Am J Clin Nutr.* **94,** 1600-6.
455 49. D'Angelo S, Gormley IC, McNulty BA, *et al.* (2019) Combining biomarker and food intake
456 data: calibration equations for citrus intake. *Am J Clin Nutr.* **110,** 977-983
457 50. Hu FB. (2002) Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin*
458 *Lipidol.* **13,** 3-9.
459 51. Andersen MB, Rinnan Å, Manach C, *et al.* (2014) Untargeted metabolomics as a screening tool
460 for estimating compliance to a dietary pattern. *J Proteome Res.* **13,** 1405-18.
461 52. Acar E, Gürdeniz G, Khakimov B, *et al.* (2019) Biomarkers of Individual Foods, and Separation
462 of Diets Using Untargeted LC-MS-based Plasma Metabolomics in a Randomized Controlled
463 Trial. *Mol Nutr Food Res.* **63,** doi: 10.1002/mnfr.201800215.
464 53. Esko T, Hirschhorn JN, Feldman HA, *et al.* (2017) Metabolomic profiles as reliable biomarkers
465 of dietary composition. *Am J Clin Nutr.* **105,** 547-54.
466 54. Rebholz CM, Lichtenstein AH, Zheng Z, *et al.* (2018) Serum untargeted metabolomic profile
467 of the Dietary Approaches to Stop Hypertension (DASH) dietary pattern. *Am J Clin Nutr.* **108,**
468 243-55.
469 55. Garcia-Perez I, Posma JM, Gibson R, *et al.* (2017) Objective assessment of dietary patterns by
470 use of metabolic phenotyping: a randomised, controlled, crossover trial. *Lancet Diabetes*
471 *Endocrinol.* **5**, 184-95.

56. Lindqvist HM, Rådjursöga M, Malmodin D, *et al.* (2019) Serum metabolite profiles of habitual diet: evaluation by 1H-nuclear magnetic resonance analysis. *Am J Clin Nutr.* **110,** 53-62.
57. Macias S, Kirma J, Yilmaz A, *et al.* (2019) Application of 1H-NMR Metabolomics for the Discovery of Blood Plasma Biomarkers of a Mediterranean Diet. *Metabolites.* **9,** 201-215
58. McCullough ML, Maliniak ML, Stevens VL, *et al.* (2019) Metabolomic markers of healthy dietary patterns in US postmenopausal women. *Am J Clin Nutr.* **109,** 1439-51.
59. Tong TYN, Koulman A, Griffin JL, *et al.* (2019) A Combination of Metabolites Predicts Adherence to the Mediterranean Diet Pattern and Its Associations with Insulin Sensitivity and Lipid Homeostasis in the General Population: The Fenland Study, United Kingdom. *J Nutr.* doi.org/10.1093/jn/nxz263
60. Gibbons H, Carr E, McNulty BA, *et al.* (2017) Metabolomic-based identification of clusters that reflect dietary patterns. *Mol Nutr Food Res.* **61,** doi.org/10.1002/mnfr.201601050
61. Johnson RK, Vanderlinden L, DeFelice BC, *et al.* (2019) Metabolite-related dietary patterns and the development of islet autoimmunity. *Sci Rep.* **9,** 14819.

487 **Table 1. Outline of approaches used to classify consumers of specific foods using food intake biomarkers**.

Consumer group	Biofluid	Dietary Assessment	Classification method	Biomarkers	Reference
Orange juice	24 hour urine	24 hour dietary recall	ROC AUC sensitivity & specificity= 92.3% & 80.6% (Test set citrus fruit consumers vs non-consumers)	Proline betaine	Heinzmann *et al.* [30]
Plant foods	24 hour urine	3DFD	Comparison plots of reported intake vs peak areas (> 20th percentile) for consumers & (</> 80th percentile) non-consumers of specific plant foods	Cabbage: iberin N-acetyl cysteine Beetroot: 4-ethyl-5-methylamino-pyrocatechol sulphate Strawberry: 2,5-dimethyl-4-methoxy-3(2H)-furanone sulphate Walnut: 5-hydroxyindole-3-acetic acid	Andersen *et al.* [36]
Coffee	Morning spot urine	24 hour dietary recall	ROC AUC = 0.95 (High vs non consumers)	Atractyligenin glucuronide	Rothwell *et al.* [32]
Banana	24 hour urine	24 hour dietary recall	ROC AUC = 0.9 (Average AUC of high vs non- & low vs non-consumers)	Methoxyeugenol glucuronide & dopamine sulfate	Vazquez-Manjarrez *et al.* [39]
Sugar-sweetened beverages	Fasting first void	4DFD	ROC AUC = 0.8 (Consumers vs non-consumers)	Citrulline, Formate, Isocitrate, Taurine	Gibbons *et al.* [40]
Beer	Multiple postprandial urine samples	Actual Intake	ROC AUC = 1 (Before & after beer intake)	(Sum of isocohumulone, isoad/ humulones, tricyclocohumol & tricyclohumol), NMT sulfate, pGlu-pro, & 2-ethyl malate	Gurdeniz *et al.* [41]

Consumer group	Biofluid	Dietary Assessment	Classification method	Biomarkers	Reference
Wine	Baseline spot urine	137-item FFQ	ROC AUC = 92.4% (Consumers vs non-consumers)	Tartrate & ethyl glucuronide	Vazquez-Fresno *et al.* [42]
Whole grain	Fasting blood samples	N/A	% misclassification rate = 9-2.1 Agreement: Cohen's weighted kappa statistic = 0.238 = slight/fair classification (high vs low consumers, quartiles)	Total plasma ARs	Ross *et al.* [37]
Whole grain Bread	Baseline spot urine	137-item FFQ	ROC AUC= 93.1% for positive mode ROC AUC= 93.7% for negative mode (Whole-grain bread vs non-bread consumers)	HHPAA, HPPA, HMBOA, 3-ICA, Enterolactone, Pyrraline, Riboflavin DHPPA, DHPPTA, HMBOA, Pyrraline, 3-ferulic acid, dihydroferulic acid, Enterolactone	Garcia-Aloy *et al.* [43]
Cocoa	Baseline spot urine	137-item FFQ	ROC AUC= 92.6% (Consumers vs non-consumers)	AMMU, DHPV glucronide & sulphate, 3- & 7-methylxanthine, 3-Methyluric acid, 3,7-Dimethyluric acid, Theobromine, MHPV	Garcia – Aloy *et al.* [44]
42 foods & food groups	Nonfasted serum samples	153-item FFQ	All ROC AUC values ≥0.65 (High vs low consumers, quintiles)	199 total metabolites identified. 43 metabolites were most discriminative for each food group.	Wang *et al.* [38]

ROC, receiver operating characteristic; AUC, area under the curve; vs, versus; 3DFD, 3 day food diary; UPLC-qTOF-MS, ultra-high performance liquid chromatography quadrupole time of flight mass spectrometry; 4DFD, 4 day food diary; NMT, N-methyl tyramine; pGlu-pro, Pyro-glutamyl proline; FFQ, food frequency questionnaire; N/A, not applicable; ARs, alkylresorcinols; HHPAA, 2-hydroxy-N-(2-hydroxyphenyl) acetamide; HPPA, 2-hydroxy-N-(2-hydroxyphenyl) acetamide; HMBOA, 2-hydroxy-7-methoxy-2H-1,4-benzoxazin-3-one; 3-ICA, 3-indolcarboxylic acid glucuronide; DHPPA, 3-(3,5-dihydroxyphenyl) propanoic acid; DHPPTA, 5-(3,5-dihydroxyphenyl) pentanoic acid; AMMU, 6-amino-5[N-methylformylamino]-1-methyluracil; MHPV, Methoxyhydroxyphenylvalerolactone; DHPV, 5-(3',4'-dihydroxyphenyl)-valerolactone.

Table 2. Summary of studies using food intake biomarkers to quantify intake.

Food/ Food group	Biofluid	Dietary Assessment	Quantification method	Performance Measurement	Biomarkers	Reference
Citrus	Fasting first void urine	4DFD	Calibration curve	Bland Altman (bias = 4.3g)	Proline Betaine	Gibbons *et al.* [45]
Chicken	Fasting first void urine	4DFD	Calibration curve	Bland Altman	Guanidoacetate	Yin *et al.* [46]
Grapes	24 hour urine	Actual Intake	Calibration curve	Correlation coefficient r^2= 0.9	Tartaric acid	Garcia-Perez *et al.* [47]

4DFD, 4 day food diary.

498 **Table 3. Using food intake biomarkers for of the study of dietary patterns**

Dietary Pattern	Classification method	Performance analysis	Validation population	Biomarkers	Reference
New Nordic vs Average Danish Diet	PLS-DA between two diets	Misclassification rate for two dietary patterns in validation set (N=139) =19%	Randomly selected samples from training set	Selected 67 metabolite markers of individual foods.	Andersen *et al.* [51]
New Nordic vs Average Danish Diet	PLS-DA between two diets	ROC AUC positive mode = 0.88 negative mode = 0.74	Test set: same population as training set (30%:70%)	NND = pipecolic acid betaine (whole grain), TMAO, & prolyl hydroxyproline (fish intake), higher PUFA PCs ADD = theobromine (chocolate) & proline betaine (citrus), amino acid metabolites (indolelactic acid & hydroxy-3-methylbutyrate) & fat metabolites (butyryl carnitine).	Acar *et al.* [52]
Low fat, low CHO & low GI diets	Bayesian network classification models	95% of withheld data classified correctly	Same samples as used to build the model	Identified 152 differential metabolites including DAGs & TAGSs, BCAAs, & markers reflecting metabolic status	Esko *et al.* [53]
Healthy & Unhealthy diets	Two-step cluster analysis	94% of validation population correctly classified	Separate healthy eating intervention population (NutriTech Study N= 49)	Healthy cluster had higher levels of hippurate, betaine, anserine, N-phenylacetylglutamine, 3-hydroxybutyrate, citrate, tryptophan & 2-aminoadipate.	Gibbons *et al.* [60]

Dietary Pattern	Classification method	Performance analysis	Validation population	Biomarkers	Reference
				Unhealthy cluster had higher levels of creatinine, glycylproline, N-acetylglutamate & theophylline	
Four diets: variable adherence to WHO guidelines	MCCV–PLS-DA	Used models based on diets 1 and 4 urinary profiles to predict consumption of diets 2 and 3 (Skilling's-Mack test $p = 7.21x10^{-9}$). Significant associations between diet scores and urinary metabolite profiles in external validation populations ($p < 0.0001$ for both)	Internal validation and two separate external validation populations (INTERMAP UK Cohort, N=225 & a Danish cohort, N=66)	Specific metabolites known to be associated with healthy eating foods: hippurate (F&V), 4-hydroxyhippurate (fruits), & S-methyl-L-cysteine-sulfoxide (cruciferous vegetables)	Garcia-Perez *et al.* [55]
DASH diet F&V diet Control diet	PLS-DA between DASH & each of other two diets	C statistic = 0.961 between DASH diet & Control diet	Test set: same population as training set (33%: 66%)	10 most influential metabolites: N-methylproline, stachydrine, tryptophan betaine, theobromine, 7-methylurate, chiroinositol, 3-methylxanthine, methyl glucopyranoside (α & β), β-cryptoxanthin, & 7-methylxanthine	Rebholz *et al.* [54]
aMED AHEI-2010 DASH HEI-2015	OPLS-DA between highest (Q5) & lowest (Q1) quintile	ROC AUC for top 10 most discriminating metabolites between Q5 & Q1: aMED = 0.77	Test set from the same population as training set (50%:50%).	aMED: 2 sphingomyelins, hydroxy-CMPF, DHA, EPA, γ- & β-tocopherol	McCullough *et al.* [58]

Dietary Pattern	Classification method	Performance analysis	Validation population	Biomarkers	Reference
	for each dietary pattern score	AHEI-2010 = 0.86 DASH = 0.86 HEI-2015 = 0.76		AHEI-2010: Hydroxy-CMPF, CMPF, DHA, sphingomyelin, EPA, carotene diol DASH: β-cryptoxanthin, sphingomyelin, γ- & β-tocopherol, galactonate, hydroxy-CMPF HEI-2015: DHA, EPA, hydroxy-CMPF, carotene diol, β-cryptoxanthin, ergothioneine	
Meat eating & avoidance	OPLS-DA between meat eaters versus non-eaters & vegans vs non-vegans	ROC AUC; Meat eaters vs non-eaters = 1 (Classified 97.5% correctly) Vegans vs non-vegans = 0.98 (Classified 92.5% correctly)	Classification of diet in same population	Serum metabolites higher in meat-eaters & non-vegans: branched chain amino acids, 3-hydroxyisobutyrate & lysine. Higher in vegans & non-meat eaters: creatine, glycine, glutamate, trimethylamine & 2-aminobutyrate.	Lindqvist *et al.* [56]
Med Diet	PLS-DA between low & high MDS	ROC AUC of citric acid & pyruvate = 0.74	Did not validate	Top five discriminative metabolites: Citric acid, myo-inositol, pyruvic acid, mannose & betaine	Macias *et al.* [57]

Dietary Pattern	Classification method	Performance analysis	Validation population	Biomarkers	Reference
Med Diet	Backwards stepwise regression between metabolite score & MDS	Spearman's correlation between metabolite score & MDS ($p = 0.42$)	Test set from the same population as training set (50%:50%).	Nuts, cereals & red/ processed meat contributed to acylcarnitines Fruit intake & amino acids/ amines Fish intakes & phospholipid concentrations	Tong *et al.* [59]
De novo infant pre-T1D infant dietary patterns	Reduced regression analysis	NA	NA	Dietary pattern 1: PC (34:2) Dietary pattern 2: SM (d41:2), GlcCer (d41:1) & PC (p-32:0) or PC (o-32:1). Dietary pattern 3: (protective for a type 1 diabetes related autoantibody response), PC (34:3) & PC (p-32:0) or PC (o-32:1) & lower concentrations of SM (d41:2)	Johnson *et al.* [61]

PLS-DA, partial least squares discriminant analysis; ADD, average Danish diet; NND, new Nordic diet; ROC, receiver operating characteristics; AUC, area under the curve; TMAO, trimethylamine oxide; PUFA, polyunsaturated; PCs, phosphotidylcholines; CHO, carbohydrate; GI, glycaemic index; DAGs, diacylglycerols; TAGs, triacylglycerols; BCAAs, branched chain amino acids; DASH, Dietary advice to stop hypertension; WHO, World Health Organisation; MCCV, Monte Carlo cross validation; F&V, fruit and vegetables; aMED, alternate Mediterranean diet score; AHEI-2010, alternate Healthy Eating Index; HEI-2015, Healthy Eating Index; OPLS-DA, orthogonal partial least squared discriminant analysis; CMPF, 3-carboxy-4-methyl-5-propyl-2-furanpropanoate; DHA, docosahexaenoic acid; EPA, eicosapentaenoic acid; Med Diet, Mediterranean; MDS, Mediterranean Diet Score; NA, not applicable; PC, phosphatidylcholine; SM, sphingomyelin; GlcCer, glucosylceramides.

VALIDATION OF FOOD INTAKE BIOMARKERS

Plausibility
Is there a chemical or biological reason for increase in biomarker concentration?

Time-response
Is there a kinetic response?
Do repeated measures show same response?

Dose-Response
As the food portion increases does biomarker concentration also increase?

Robustness
Are there other confounding foods?
Is biomarker identified after a complex meal?

Reliability
How does biomarker intake estimation compare to other dietary assessment methods?

Stability
Is there any degradation of the biomarker in the biofluid?

Performance
Has lab analysis examined biomarker accuracy, precision, sensitivity & specificity?

Reproducibility
Are the results reproducible in other labs and within other populations?