

Aberystwyth University

Challenges Associated With the Design and Deployment of Food Intake Urine Biomarker Technology for Assessment of Habitual Diet in Free-Living Individuals and Populations

Beckmann, Manfred; Wilson, Thomas; Lloyd, Amanda; Torres, Duarte P.M.; Goios, Ana C.L.; Willis, Naomi D.; Lyons, Laura; Phillips, Helen; Mathers, John M.; Draper, John

Published in:
Frontiers in Nutrition

DOI:
[10.3389/fnut.2020.602515](https://doi.org/10.3389/fnut.2020.602515)

Publication date:
2020

Citation for published version (APA):
Beckmann, M., Wilson, T., Lloyd, A., Torres, D. P. M., Goios, A. C. L., Willis, N. D., Lyons, L., Phillips, H., Mathers, J. M., & Draper, J. (2020). Challenges Associated With the Design and Deployment of Food Intake Urine Biomarker Technology for Assessment of Habitual Diet in Free-Living Individuals and Populations: A Perspective. *Frontiers in Nutrition*, 7, [602515]. <https://doi.org/10.3389/fnut.2020.602515>

Document License CC BY

General rights

Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400
email: is@aber.ac.uk



Challenges Associated With the Design and Deployment of Food Intake Urine Biomarker Technology for Assessment of Habitual Diet in Free-Living Individuals and Populations—A Perspective

Manfred Beckmann¹, Thomas Wilson^{1†}, Amanda J. Lloyd^{1†}, Duarte Torres^{2,3}, Ana Goios^{2,3}, Naomi D. Willis⁴, Laura Lyons¹, Helen Phillips¹, John C. Mathers⁴ and John Draper^{1*}

OPEN ACCESS

Edited by:

Alessandra Bordoni,
University of Bologna, Italy

Reviewed by:

Fulvio Mattivi,
University of Trento, Italy
Diego A. Moreno,
Spanish National Research
Council, Spain

*Correspondence:

John Draper
jhd@aber.ac.uk

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Nutritional Epidemiology,
a section of the journal
Frontiers in Nutrition

Received: 03 September 2020

Accepted: 29 October 2020

Published: 25 November 2020

Citation:

Beckmann M, Wilson T, Lloyd AJ, Torres D, Goios A, Willis ND, Lyons L, Phillips H, Mathers JC and Draper J (2020) Challenges Associated With the Design and Deployment of Food Intake Urine Biomarker Technology for Assessment of Habitual Diet in Free-Living Individuals and Populations—A Perspective. *Front. Nutr.* 7:602515. doi: 10.3389/fnut.2020.602515

¹ Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, United Kingdom, ² Faculty of Nutrition and Food Sciences, University of Porto, Porto, Portugal, ³ Epidemiology Research Unit (EPIUnit), Institute of Public Health, University of Porto, Porto, Portugal, ⁴ Human Nutrition Research Centre, Population Health Sciences Institute, William Leech Building, Newcastle University, Newcastle-upon-Tyne, United Kingdom

Improvement of diet at the population level is a cornerstone of national and international strategies for reducing chronic disease burden. A critical challenge in generating robust data on habitual dietary intake is accurate exposure assessment. Self-reporting instruments (e.g., food frequency questionnaires, dietary recall) are subject to reporting bias and serving size perceptions, while weighed dietary assessments are unfeasible in large-scale studies. However, secondary metabolites derived from individual foods/food groups and present in urine provide an opportunity to develop potential biomarkers of food intake (BFIs). Habitual dietary intake assessment in population surveys using biomarkers presents several challenges, including the need to develop affordable biofluid collection methods, acceptable to participants that allow collection of informative samples. Monitoring diet comprehensively using biomarkers requires analytical methods to quantify the structurally diverse mixture of target biomarkers, at a range of concentrations within urine. The present article provides a perspective on the challenges associated with the development of urine biomarker technology for monitoring diet exposure in free-living individuals with a view to its future deployment in “real world” situations. An observational study ($n = 95$), as part of a national survey on eating habits, provided an opportunity to explore biomarker measurement in a free-living population. In a second food intervention study ($n = 15$), individuals consumed a wide range of foods as a series of menus designed specifically to achieve exposure reflecting a diversity of foods commonly consumed in the UK, emulating normal eating patterns. First Morning Void urines were shown to be suitable samples for biomarker measurement. Triple quadrupole mass spectrometry, coupled with liquid chromatography, was used to assess simultaneously the behavior of a panel of 54 potential BFIs. This panel of chemically diverse biomarkers, reporting intake of a wide range of commonly-consumed foods, can be extended successfully as new biomarker leads are discovered. Towards

validation, we demonstrate excellent discrimination of eating patterns and quantitative relationships between biomarker concentrations in urine and the intake of several foods. In conclusion, we believe that the integration of information from BFI technology and dietary self-reporting tools will expedite research on the complex interactions between dietary choices and health.

Keywords: dietary intake, metabolomics, biomarker of food intake (BFI), urinary biomarkers, habitual diet

INTRODUCTION

There is a rich history of nutrition research spanning many decades, much of which has had at its core a need for accurate information on dietary intake for investigation of the links between exposure to individual food/food groups and specific health outcomes. Food intervention projects commonly rely on participants collecting pre-prepared foods from research centres for consumption at home and then confirming compliance at a later date (1, 2). On the other hand, large-scale nutritional epidemiological projects and nutrition surveys involving free-living individuals consuming their habitual diet rely almost totally on self-reporting of dietary exposure. Long-established tools to collect self-reported quantitative dietary information include Food Frequency Questionnaires (FFQs), diet diaries, and dietary recall methodology (3). However, because of the complexity of eating patterns and the conceptual and practical difficulties in recording or recalling the types and amounts of foods and beverages consumed, errors in self-reporting of dietary intakes by cognitively-able individuals is commonplace and substantial (4, 5) and can be exacerbated in those who are overweight or obese (6, 7).

Secondary metabolites derived from individual foods or food groups present in human biofluids can provide potential biomarkers of food intake, for reviews see (8–18). The inclusion of biomarker technology in dietary assessment could help to overcome some of the limitations of traditional dietary methodologies by providing additional objective estimates of food exposure (19). Unlike blood, urine is easy to collect and it provides an integrated estimate of exposure over several hours. For a panel of dietary biomarkers to have any significant utility, it is essential that its coverage is as comprehensive as

possible. Using data-driven approaches, we have shown that the potential utility of a biomarker is dependent on the type, portion size, and frequency of consumption of individual foods (20). Data concerning nationally-representative estimates of intakes of foods by the UK population are collected by the UK National Diet and Nutrition Survey (NDNS) (21) and this database can be explored to identify foods and food groups for which dietary exposure biomarker discovery might be feasible and relevant (1, 2, 22).

Over the past decade, our collaborative research projects and those of other teams (see **Supporting Data 1** for a comprehensive list) have contributed to the discovery of putative dietary intake urinary biomarkers of specific foods including poultry and red meat (23–28) citrus fruits (29, 30), crucifers (31, 32), oily fish (26, 27, 32), red berries/strawberries (2, 32–34), wholegrain/rye (35–37), sugary drinks (38, 39), artificial sweeteners (2, 40), peas/beans/legumes (2, 41, 42), grapes (41, 43–45), apples (41, 46, 47), and potatoes (48). In addition, consensus guidelines for the critical assessment of candidate BFIs has been established (49). These BFI candidate guidelines have focused generally on qualifying the utility of individual BFIs for monitoring exposure to specific foods/food groups. However, because effects on health are a consequence of the whole diet, it is equally important to develop approaches to assess overall dietary exposure in nutrition surveys, epidemiological studies, and clinical trials (45).

The ideal biomarker is highly specific for one food item or food group, is not detected in the biological sample of interest when the specific food item is not ingested, and shows a distinct dose- and time-dependent response following consumption (50). Although metabolites distinctive of dietary exposure to particular foods have been described, it is not uncommon to discover subsequently that putative biomarkers are not necessarily specific for individual foods and therefore much rigour needs to be applied during validation of their utility to monitor habitual dietary intake (51). For application in the real world, the use of multi-metabolite biomarker panels may provide more reliable estimation of dietary exposure than a single-biomarker approach [reviewed by (52)]; such panels need to have comprehensive coverage and to be extendable (53). For this reason, in the future it will be important to evaluate biomarker performance in the context of complex exposures to multiple foods, with different food formulations, cooking, and processing methods and within complex meals, in eating patterns the target study population is likely to experience (2).

Optimal sampling requirements for urine biomarker analysis will be dependent largely on the study design and objectives (**Table 1**). For example, a food intervention study with free-living

Abbreviations: AUC, area under the ROC (Receiver Operator Characteristic) curve; BFI, biomarker of food intake; CRN, Clinical Research Network; FFQ, Food Frequency Questionnaire; FMV, First Morning Void; HESI, heated electrospray ionisation; HILIC, Hydrophilic Interaction Liquid Chromatography; HPLC, high-performance liquid chromatography; IAN-AF, Portuguese National Food, Nutrition and Physical Activity Survey; ISRCTN, International Standard Randomised Controlled Trials Number; LC-QQQ-MS, liquid chromatography triple quadrupole mass spectrometry; LoD, limit of detection; logP, partition coefficients; LoQ, limit of quantification; MACCS, Molecular ACCess System; MAIN, Metabolomics at Aberystwyth, Imperial and Newcastle; MDS, multi-dimensional scaling; MRC, Medical Research Council; MRM, multiple reaction monitoring; MS, Mass Spectrometry; NDNS, National Diet and Nutrition Survey; PCA, Principal Components Analysis; QC, Quality Control; RF, Random Forest; RI, refractive index; ROC, Receiver Operator Characteristic; RP, reverse phase; RSD, relative standard deviation; SG, specific gravity; SRM, Selected Reaction Monitoring; UHPLC, Ultra High Performance Liquid Chromatography.

TABLE 1 | Study objectives and biomarker of food intake (BFI) requirements.

Example study objectives	Typical sampling requirements		Biomarker requirements		Data requirements		Study example and reference
	Single sample only	Multiple samples	Biomarker(s) of only one food/food group	Comprehensive biomarker panel	Quantitative or semi-quantitative measurement	Exposure range assignment	
A Confirmation of participant compliance in a food intervention study or validation of a proposed biomarker focusing on a single food/food group, short term or long term		Y	Y		Y		A validation trial: (54) A compliance trial: (55, 56)
B Biomarker discovery and/or validation in a free-living population following a meal plan emulating normal eating patterns		Y		Y	Y		MAIN study: (1, 2, 22)
C Investigation of individual "metabotype" in relation to interaction with specific dietary chemicals	?	?	Y		Y		Food4me study: (57)
D Assessment of habitual (e.g., weekly, monthly, and annual) eating behavior of individuals		Y		Y	?	?	PREDIMED trial: (58, 59)
E Observational epidemiological survey of eating habits in a large population	Y			Y	Y	?	IAN-AF: (60); EPIC: (27, 61)
F Cohort stratification by dietary exposure levels to specific foods/food groups in a small clinical trial		Y		Y		Y	MAIN study: (45)

participants lasting several weeks investigating links between a health outcome and a specific food/food group will require appropriate samples on multiple days taken at random to assess compliance with dietary intake targets (1, 2). In contrast, assessing the general eating habits of a large population in an epidemiological survey may only require sampling of a large number of people on a single random day or multiple days (60). Any urine sampling procedure would need to be (i) acceptable for volunteers to provide samples repeatedly, (ii) require minimal researcher time and cost, and (iii) deliver samples with high quality information content. The theoretical optimal types of urine(s) to be sampled [e.g., spot, cumulative (i.e., "phase" of day) or 24 h] will also depend on study objectives (62) and, in many instances, the sampling strategy will be limited by cost constraints or the practicalities of collection. Twenty-four hour urine samples and single spot urine samples taken at random times during the day are commonly collected to monitor discrete aspects of human physiology, metabolism, or "exposome" in clinical trials and surveys (63–65). Unfortunately, such samples provide information only in relation to very recent eating behaviour and may be of limited utility in nutrition studies where the focus is on the whole diet or on the intakes of foods that are not eaten frequently. Additionally, eating behaviour and hydration levels can be very different between individuals

in free-living populations and the fact that excretion half-lives of specific metabolites can vary enormously (49), means that research protocols must be in place to manage adequately these sources of variability in any biomarker discovery and validation strategy (62, 66).

Where the aim is to estimate absolute intake of specific foods, or the frequency of exposure, it will be desirable to generate quantitative or semi-quantitative data on BFI concentrations in urine. However, for other studies, it may be sufficient to be able to assign each individual into an exposure range (e.g., high-medium-low), typical of a specific reference population. Urine collection(s), sample processing, and the analytical methodology can be optimised for a target metabolite when using a single biomarker to monitor exposure to a single food/food group. In contrast, the desire to monitor habitual diet comprehensively using a panel of biomarkers requires the analytical approach to manage the complex physio-chemical attributes of the diverse range of putative biomarkers currently described, as well as coping with metabolites exhibiting differential stability during collection, transport, and storage. Other practical issues such as the commercial availability, costs, solubility, and stability of pure chemicals in mixtures as quantitation standards will also impact on the design of analytical solution likely to offer

scope for simultaneous measurement of a large number of metabolite targets.

The assessment of eating behaviour in free-living individuals is important in a wide range of types of nutrition research (Table 1), ranging from clinical trials investigating the mode of action of potentially beneficial “bioactive” compounds in individuals, to general surveys of national eating habits in large populations. Although considerable effort is being expended on BFI discovery and validation, there is an equally urgent need to consider the future challenges for effective deployment of dietary intake biomarker technology to assess habitual diet within populations. To summarise, these major challenges include:

- Strategies for validation of food intake biomarkers suitable for assessment of habitual dietary exposure;
- Standardised urine sampling approaches, including collection, temporary storage, transport, and long term biobanking;
- Development of biomarker analytical methodology, using a multi-panel of biomarkers, that is able to integrate new markers as they become validated;
- Algorithms to convert raw biomarker data into meaningful estimates of food intake and/or overall diet quality.

The present article aims to provide a perspective on *some* of these challenges associated with the development of urine biomarker technology to monitor recent or habitual dietary intake in free-living individuals with a view to its future deployment in “real world” situations. For example the UK government’s “Better Health” campaign, a 12 week fitness and healthy eating plan announced in July 2020 to help Britons lose weight and reduce their risk of serious complications should they contract COVID-19 (<https://www.nhs.uk/better-health/>). One of the key aims of the plan is to encourage people to make healthier food choices. We believe that the BFI technology we have been developing over the past 10 years, alongside the low-effort, minimally intrusive urine sampling strategies (62, 67) will soon be validated to the point that they can be used to reproducibly and objectively monitor the effectiveness of such plans at a population level. A workflow summarising the overall experimental strategy is illustrated in Figure 1.

MATERIALS AND METHODS

Ethics Approval and Consent to Participate

For *Study 1*, ethical approval was obtained from the National Commission for Data Protection, the Ethical Committee of the Institute of Public Health of the University of Porto and from the Ethical Commissions of each one of the Regional Administrations of Health. All participants gave written informed consent, and the study was carried out in accordance with the Declaration of Helsinki. The MAIN (Metabolomics at Aberystwyth, Imperial, and Newcastle) food intervention trial at Newcastle (*Study 2*) was approved by the East Midlands—Nottingham 1 National Research Ethics Committee (14/EM/0040). Caldicott approval for storage of data and data protection was granted by Newcastle-upon-Tyne Hospitals NHS Foundation Trust [6896(3109)]. The MAIN food intervention trial in Newcastle was adopted into the UK Clinical

Research Network (CRN) Portfolio (16037) and is registered with International Standard Randomised Controlled Trials Number (ISRCTN), 88921234.

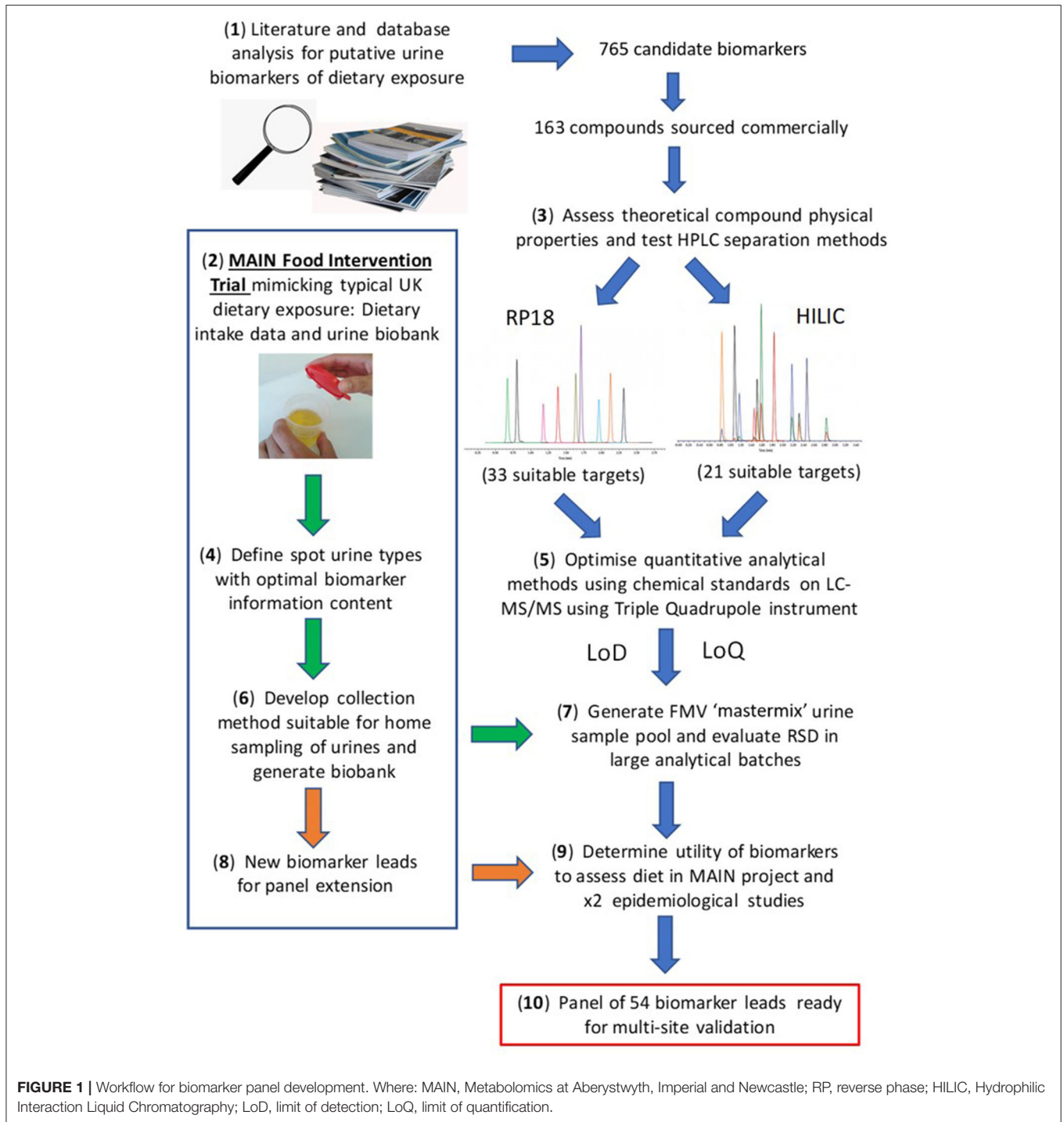
The participants provided written informed consent to participate in each study, taken by an appropriately trained researcher. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Epidemiological Study and Urine Sampling

Study 1 involved community-living individuals consuming a freely-chosen diet. The participants ($n = 95$) were volunteers who participated in the Portuguese National Food, Nutrition, and Physical Activity Survey (IAN-AF), whose aims and methods have been described previously (60). A 24 h dietary record was collected by trained nutritionists using the “eAT24” Software (68) which facilitates the assessment of dietary data using an automated multiple-pass method (5 steps) (69). Participants were asked to collect urine samples on the day before the second 24 h dietary record. Urine samples were collected in two separate containers. The first one (a 2,700 mL container identified as container A) was used to collect all urine passed during the day before the interview, except the first void of that morning. A second one (a 500 mL container identified as container B) was used to collect just the First Morning Void (FMV) on the day of the second interview. No preservatives were added to the urine containers, and the participants were asked to keep the samples refrigerated (4°C) throughout the collection period. Participants were asked to fill in a questionnaire with the time of the beginning and the end of collections, details of any medication, and whether or not they had any problems or missed urine collections. At the laboratory, urine samples were weighed and mixed. The weights of urine from containers A and B were quantified separately and a proportionally pooled 24 h urine sample (identified as “24 h urine”) was prepared by using samples A and B. From each participant, both urine samples were aliquoted: 1 × 45 mL (in 50 mL Falcon pre-labelled tube) + 10 × 1.5 mL (in 2 mL pre-labelled microtubes). These aliquots were refrigerated immediately before being moved to −80°C storage, within 24 h, for further analysis.

Food Intervention Study Design and Urine Sampling

The MAIN project at Newcastle included two controlled food intervention studies in free-living people who consumed the test foods as part of two 3day menu plans, designed to generate six distinctive “Menu Days” (1, 22). Participants were provided with all the foods and ingredients to prepare and consume meals at home, following the prescribed menus. Within this manuscript (*Study 2*) we have used data from 15 of the individuals from the second 3-day menu plan (8 male, 7 female; non-smokers; age: 21–74). We implemented urine sampling methods based on our previous studies (23, 70) and asked participants to collect a series of urine samples including the FMV the day after each menu plan. Participants collected urine samples in a plastic jug



and transferred aliquots into labelled sterile 25 mL Universal tubes. Six of these 15 participants (2 female, non-smokers, age range 22–59) also collected FMV urine samples at home using the vacuum transfer system (67). All samples were placed in an opaque cool bag and stored at home in a fridge at 4°C for up to 4 days and then brought to the research facility in Newcastle at the end of the study week. Universal tubes were stored immediately at –80°C and the vacuum tubes remained at

4°C for a further 2 weeks before storage at –80°C. Samples were then transported to the analytical facility in Aberystwyth on dry ice for metabolite analysis.

Urine Sample Preparation

Urine samples were prepared and adjusted as reported previously (1, 22). In brief, all urine samples were normalised by refractive index (RI) prior to analysis to account for differences in fluid

intake by participants and to ensure that all Mass Spectrometry (MS) measurements were made within a similar dynamic range within the linear range of the instrument. Samples were defrosted overnight at 4°C, centrifuged (1,600 × g for 5 mins at 4°C), placed on ice and aliquots of thawed urine (1,000 µL) were transferred into labelled 2 mL Eppendorf tubes. The remaining sample were returned to a -20°C freezer. An OPTI Hand Held Refractometer (Bellingham Stanley™ Brix 54 Model) was used to record the specific gravity (SG). Using these data, aliquots of the required amounts of urine from centrifuged 2 ml Eppendorf tubes and ultra-pure (18.2 Ω) H₂O were transferred into new tubes for extraction; this ensured that all samples had the same RI.

Strategy for Selection of Candidate Dietary Exposure Biomarkers

The selection of biomarkers was initiated with a literature search to generate an initial “long list” of food-related metabolites with potential for inclusion in a panel of biomarkers that would provide comprehensive coverage of food items consumed in the MAIN Study (see **Supporting Data 2** for a summary of the foods). The search was carried out using Google Scholar and Web of Knowledge using the following search terms in a range of combinations “biomarkers + urine + food + dietary + BFI” and ended on 22/06/2020. Publications were screened and information was added to the database if they contained data relating to potential dietary exposure biomarkers measured in urine samples (see **Supporting Data 3**). Specific details on metabolite excretion profile were recorded and the availability of a commercial supply of a pure chemical standard was investigated (see **Supporting Data 3**).

Evaluation of Chemical Diversity of Biomarkers

Biomarkers were assigned to chemical class and superclass using the ClassyFire application (71). Classifications for each biomarker were retrieved using the R package classyfireR Version 0.3.3. Pairwise similarity of biomarkers was measured using the Tanimoto Distance after converting structural representations of each biomarker to its MACCS (Molecular ACCess System) fingerprint. Fingerprints were generated using the get.fingerprint function from the R package rcdk (Version 3.5) and distances computed using the fp.sim.matrix function from the R package fingerprint. The resultant matrix of fingerprint distances, was then reduced to two dimensions using the cmdscale function. Chemical descriptors (-logP) were calculated using the rcdk package.

Sample Analysis by Liquid Chromatography Triple Quadrupole Mass Spectrometry (LC-QQQ-MS)

Methanol (primer trace analysis grade, Fisher Scientific, UK) was used for urine extraction and standards preparation. Acetonitrile (Optima® LC-MS grade, Fisher Scientific, UK), methanol (HPLC grade, Fisher Scientific, UK), and Ammonium acetate (Optima® LC-MS grade, Fisher Scientific, Belgium) were used for preparing the LC mobile phase. Water was ultra-pure water

(18.2 Ω) drawn from an Elga Purelab® flex water purifier system (Taiwan). The suppliers of chemical standards are given in **Supporting Data 4**.

Sample analysis was performed on a TSQ Quantum Ultra EMR QQQ mass spectrometer (Thermo Scientific) equipped with a heated electrospray ionisation (HESI) source. Samples were delivered using an Accela ultra-high-performance liquid chromatography (UHPLC) system (Thermo Scientific) consisting of autosampler, column heater, and quaternary UHPLC-pump. For HILIC (Hydrophilic Interaction Liquid Chromatography) analysis, chromatographic separation was performed on a ZIC-pHILIC (polymeric 5 µm, 150 × 4.6 mm) column (Merck). The mobile phase consisted of 10 mM ammonium acetate in water: acetonitrile (95:5) (A) and 10 mM ammonium acetate in water: acetonitrile (5:95) (B). The gradient program used was as follows: 0 min, 95% B (400 µL min⁻¹); 15 min, 20% B (400 µL min⁻¹); 15.01 min, 20% B (500 µL min⁻¹); 20 min, 20 % B (500 µL min⁻¹); 20.01 min, 95 % B (500 µL min⁻¹); 25 min, 95% B (500 µL min⁻¹). The HPLC was carried out in low pressure (~0–7,000 psi) operating mode with 0 psi and 650 psi as minimum and maximum pressures, respectively. For Reverse Phase (RP) analysis, chromatographic separation was performed on a Hypersil Gold (1.9 µm, 200 × 2.1 mm) RP-column (Thermo Scientific). The mobile phase consisted of 0.1% formic acid in H₂O (A) and 0.1% formic acid in MeOH (B). The gradient program used was as follows: 0 min, 0% B; 0.5 min, 0% B; 5 min, 60% B; 11 min, 100% B; 13 min, 100% B; 13.01 min, 0% B; 19 min, 0% B. For RP analysis, the flow rate was maintained at 400 µL/min⁻¹. The UHPLC was carried out in high pressure (~7,000–15,000 psi) operating mode with 0 and 1,000 psi as minimum and maximum pressures, respectively. For both chromatographic analyses, column oven and autosampler tray were maintained at 60 and 14°C, respectively. To ensure consistent sample delivery, 20 µL were injected using a 20 µL loop and partial loop injection mode. After each injection, syringe, and injector were cleaned using a 10 % HPLC grade MeOH solution in ultra-pure water (1 mL flush volume; 100 µL/s⁻¹ flush speed, 2 mL wash volume) to avoid sample carryover. Mass spectra were acquired in multiple reaction monitoring (MRM) mode, in positive and negative ionisation polarities simultaneously using optimised values of collision energy and tube lens for each MRM transition (**Supporting Data 4**). Spectra were collected at a scan speed of 0.010 and 0.003 s for HILIC and RP analysis, respectively. A scan width of 0.010 m/z units and peak width (Q1, Q3) of 0.7 FWHM were used for both HILIC and RP analyses.

Raw files (ThermoFisher) were converted to mzML (72) using msconvert in the ProteoWizard tool kit (73). All further processing of mzML files was performed using the R Statistical Programming Language (74). Selected Reaction Monitoring (SRM) chromatograms were extracted from mzML files using the R library, mzR and peaks areas were calculated by extracting pre-defined chromatographic windows (based on calibration standards) around each peak apex. Absolute concentrations were calculated using a nine-point calibration curve (0.006561–100 µg mL⁻¹). The limit of detection (LoD) and limit of quantification (LoQ) of all chemical standards were calculated as the lowest

concentration of each biomarker giving a signal-to-noise ratio of 3:1 and 10:1, respectively within the linear range of each calibration curve.

Quality Control (QC) Strategy for Target Biomarkers

Reproducibility of the mixture of chemical standards was determined using the relative standard deviation (RSD) of a multi component calibration standard and an external urine QC sample using a “master mix” of pooled samples. The external urine QC sample was used to determine the effect of the resultant urine matrix on the reproducibility of selected biomarkers across multiple experiments. The external QC (as distinct from an experimental QC) allowed for longitudinal monitoring of RSD without intra-experimental bias.

Data Analysis

Principal Components Analysis (PCA) was performed using the *prcomp* function in R, with variables scaled to unit variance. Supervised classification of quantitative metabolite data was performed using Random Forest (RF) classification using the *randomForest* package (75) in R (74). For all RF models, the number of trees (*ntree*) used was 500 and the number of variables considered at each internal node (*mtry*) was the square root of the total number of variables. Accuracy, margins of classification and area under the ROC (Receiver Operator Characteristic) curve (AUC) were all used to evaluate the performance of classification models, as described previously (76). RF classification models were plotted following multi-dimensional scaling (MDS). Proximity measures for each individual observation were extracted from RF models and scaled coordinates produced using *cmdscale* on $1 - \text{proximity}$.

Spearman rank correlations of biomarker concentrations in 24 h vs. FMV urine were produced using the *rcorr* function from the R (Version 4.0.3) package *Hmisc* (Version 4.4.0). Reported *P*-values are the asymptotic *p*-values from the rank correlation. Quantile–Quantile plots were produced using the *qqnorm* function in R.

RESULTS

Selection of Target Foods and Design of a Food Intervention Study for Preliminary Survey of the Potential Utility of Urine Biomarker Technology

A major component of our strategy to develop urine biomarker technology to monitor habitual diet was the need for a biobank of urine samples from a food intervention trial that was designed to provide comprehensive exposure to foods commonly consumed in the UK. Key food groups were identified initially from *The Eatwell Guide* (77); the most commonly eaten foods were identified within each disaggregated food group using estimates of intakes of foods by the UK population from the UK National Diet and Nutrition Survey (NDNS) (21). **Supporting Data 2** describes the representative foods that were incorporated into a six-menu design as part of the MAIN food intervention

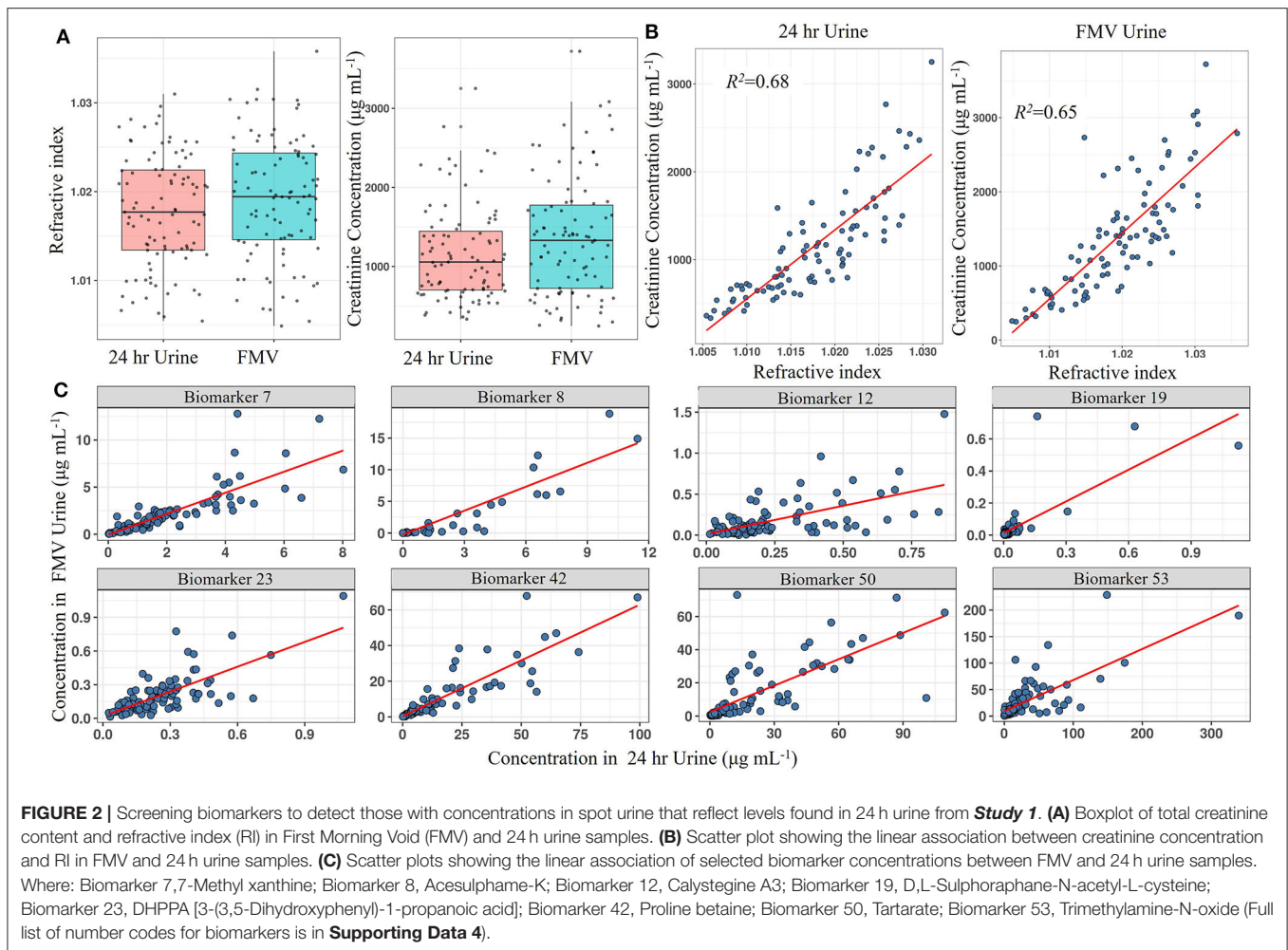
trial at Newcastle [see Lloyd et al. (1) and Willis et al. (2) for full details]. The menu plans aimed to deliver foods for BFI discovery and validation, including the assessment of BFI specificity and sensitivity within the context of the whole diet. Particularly important was consideration of the impact of the likely major sources of variance on biomarker monitoring procedures including:

- The impact of exposure to targeted foods as part of complex and mixed meals, rather than foods consumed in isolation;
- The use of average portion sizes and normal eating patterns rather than exposure to huge, unrealistic portions consumed in a fasted state;
- The impact of different food formulations, processing, and cooking methods representing the range of ways in which foods are processed and eaten;
- The dynamics of putative BFI retention in the body (to inform development of biomarkers of both acute or habitual food consumption).

The selection of biomarkers was initiated with a preliminary literature search to identify putative urinary biomarkers that would provide comprehensive coverage of each specific food/food group consumed within the six menus. This database (**Supporting Data 1**) suggested that 765 urinary metabolites were potential dietary exposure biomarker candidates as summarised in **Supporting Data 2**. It is clear from these data that there is considerable choice in terms of potential biomarkers and considerable overlap of metabolites between some foods/food groups.

Evaluation of Approaches for Urine Collection and Storage for Monitoring Habitual Dietary Exposure

Using LC-MS fingerprinting methodology, we have shown previously that the metabolome of spot urine samples taken just before bed time on the study day is compositionally very similar to the corresponding 24 h urine samples (62). In the present study, RI measurements revealed that FMV spot urine samples from a national dietary intake survey (**Study 1**) also had an almost identical overall solute concentration range to that of 24 h urines (**Figure 2A**). Since creatinine concentrations are often used as a reference for normalisation in urine samples, we evaluated the relationship between creatinine concentration and RI in both FMV spot and 24 h urines. Creatinine concentrations were within the same range in FMV spot and 24 h urines and exhibited a strong linear relationship ($R^2 = 0.65$ and 0.68 , respectively) with RI (**Figure 2B**), supporting the concept of sample normalisation to the same RI. The value of FMV spot urine samples for assessment of dietary exposure was explored further by examining the correspondence between the concentration of putative biomarkers in 24 h urine and spot urine samples using targeted, quantitative measurements of individual biomarkers. **Figure 2C** shows scatter plots of metabolite concentration in FMV urine vs. 24 h urine for eight example biomarkers from samples derived from **Study 1**. Although the actual biomarker levels varied between the two urine types there was



a strong linear relationship between concentrations in 24 h and FMV urine (**Figure 2C**). More than 50% of the biomarkers demonstrated a very strong correlation (>0.6). Further potential biomarkers exhibited a weaker correlation in concentration (**Supporting Data 4**) and it is suggested that that $r > 0.2$ (with a $p < 0.05$ from a rank correlation test) may be considered adequate. The Quantile–Quantile plot in **Supporting Data 5** shows the comparable distribution of biomarker concentrations measured in FMV and 24 h urine samples.

We have shown recently that vacuum tube technology has considerable value for spot urine sampling and that, even in the absence of preservatives, urine composition is stable for several days at 4°C (67) and under different temperature regimes. To explore further the utility of vacuum tube technology for large-scale urine sampling in community settings, we evaluated the compositional stability of FMV spot urine at 4°C for 2 weeks, to mimic longer term storage in a domestic fridge. A selection of biomarkers useful for assessment of exposure to meat, fish, wholegrain, fruit, and vegetable components of meals were targeted for analysis. Metabolite concentrations after storage in vacuum tubes at 4°C were very similar to those of the same urine samples after being frozen at –80°C (**Figure 3**).

Literature Analysis to Select Biomarker Leads for Inclusion in a Panel That Will Provide a Comprehensive Survey of Dietary Exposure

A comprehensive list of potential urinary BFIs based on a literature analysis (53) of putative dietary exposure biomarkers in various human biofluids is presented in **Supporting Data 2**. The present biomarker panel strategy aimed to assess habitual diet in individuals and populations; key to this objective was the need to use spot urine samples, specifically urine samples collected just before bedtime and FMV urines, that would be informative of overall food consumption (1, 62). A detailed examination of the dietary exposure biomarker literature was undertaken with particular emphasis on the identification of biomarker candidates present in spot urine samples > 12 h after food consumption (**Table 2** and described in further detail in **Supporting Data 3**). A shortlist of candidate biomarkers for initial biomarker panel development was generated, focusing largely on metabolites that were available from commercial providers (**Supporting Data 3**). For 28 out of 54 putative biomarkers there was already evidence in the literature of their presence in FMV urine. The majority

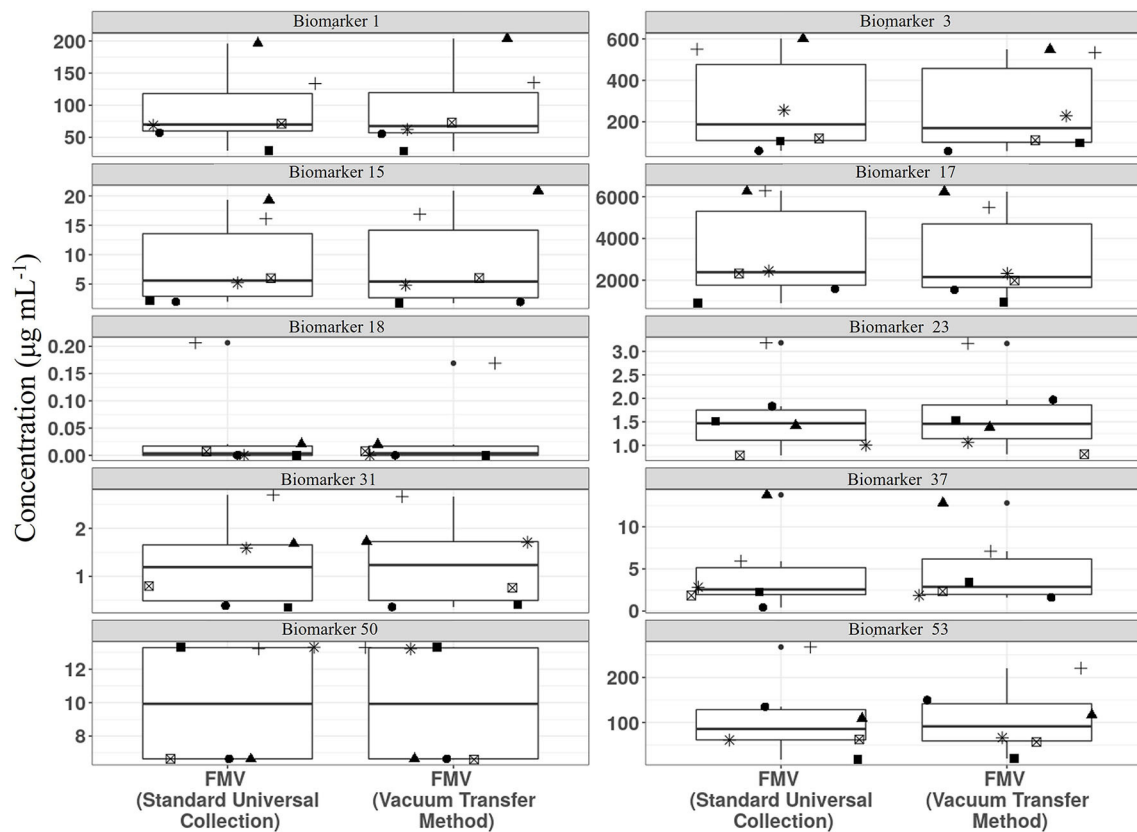


FIGURE 3 | Comparison of stability of example biomarkers in First Morning Void (FMV) urine collected in vacuum tubes stored at 4°C for 2 weeks (Vacuum transfer method) or Universal tubes (Standard Universal collection) stored at -80°C. Where: Biomarker 1, 1-Methyl histidine; Biomarker 3, 3-Methyl histidine; Biomarker 15, Carnosine; Biomarker 17, Creatinine; Biomarker 18, D,L-Sulphoraphane L-cysteine; Biomarker 23, DHPPA [3-(3,5-Dihydroxyphenyl)-1-propanoic acid]; Biomarker 31, Ferulic acid-4-O-sulphate; Biomarker 37, L-Anserine; Biomarker 50, Tartarate; Biomarker 53, Trimethylamine-N-oxide (Full list of number codes for biomarkers is in **Supporting Data 4**).

of the remaining dietary exposure biomarker leads selected had been shown to be present in 24h urine samples and so it was reasonable to expect their presence in FMV urine samples collected the day after a specific food intervention.

Examination of Biomarker Behaviour During LC-MS and Development of a Biomarker Panel Strategy

Previous analysis of published literature revealed that the great majority of dietary exposure biomarker candidates were detected and quantified using LC-MS technology (53). Chemical classification of putative biomarkers showed great structural diversity that included metabolites from 17 Chemical Classes representative of 7 Chemical Super-Classes (**Figure 4A**). *In silico* multi-dimensional scaling of structural attribute fingerprint distances shows the large diversity in chemical structure across biomarker candidates, highlighting the necessity for employing multiple chromatography systems (**Figure 4B**). Focusing specifically on partition coefficients ($\log P$) and molecular weight attributes, it is clear from the scatterplot shown

in **Figure 4C** that a large percentage of biomarker candidates were quite strongly hydrophilic. Based on this distribution a decision was made to develop an analytical strategy based largely on the use of a HILIC column to measure strongly polar chemicals and a RP (C_{18}) column to quantify less polar metabolites.

LC-QQQ-MS/MS technology is used widely for measuring, with high sensitivity, the concentration of target chemicals in complex biological samples. Quantification of pre-determined fragmentation products of targeted metabolites in expected retention time windows using MRM approaches allows the investigator to obtain data on large numbers of individual metabolites in short (10–15 min) HPLC runs. Chemical mixtures designed as calibration standards for either HILIC or RP (C_{18}) chromatography were used to optimise metabolite separation and detection conditions on a Thermo QQQ instrument (see **Supporting Data 4**). Serial dilutions of the two standard chemical mixtures (30–0.00197 $\mu\text{g ml}^{-1}$) were used to establish LoD and LoQ and to examine analytical reproducibility over several months. The reproducibility of measurement of chemical standard mixtures was determined at nine concentration levels.

TABLE 2 | Selection of biomarkers for panel development.

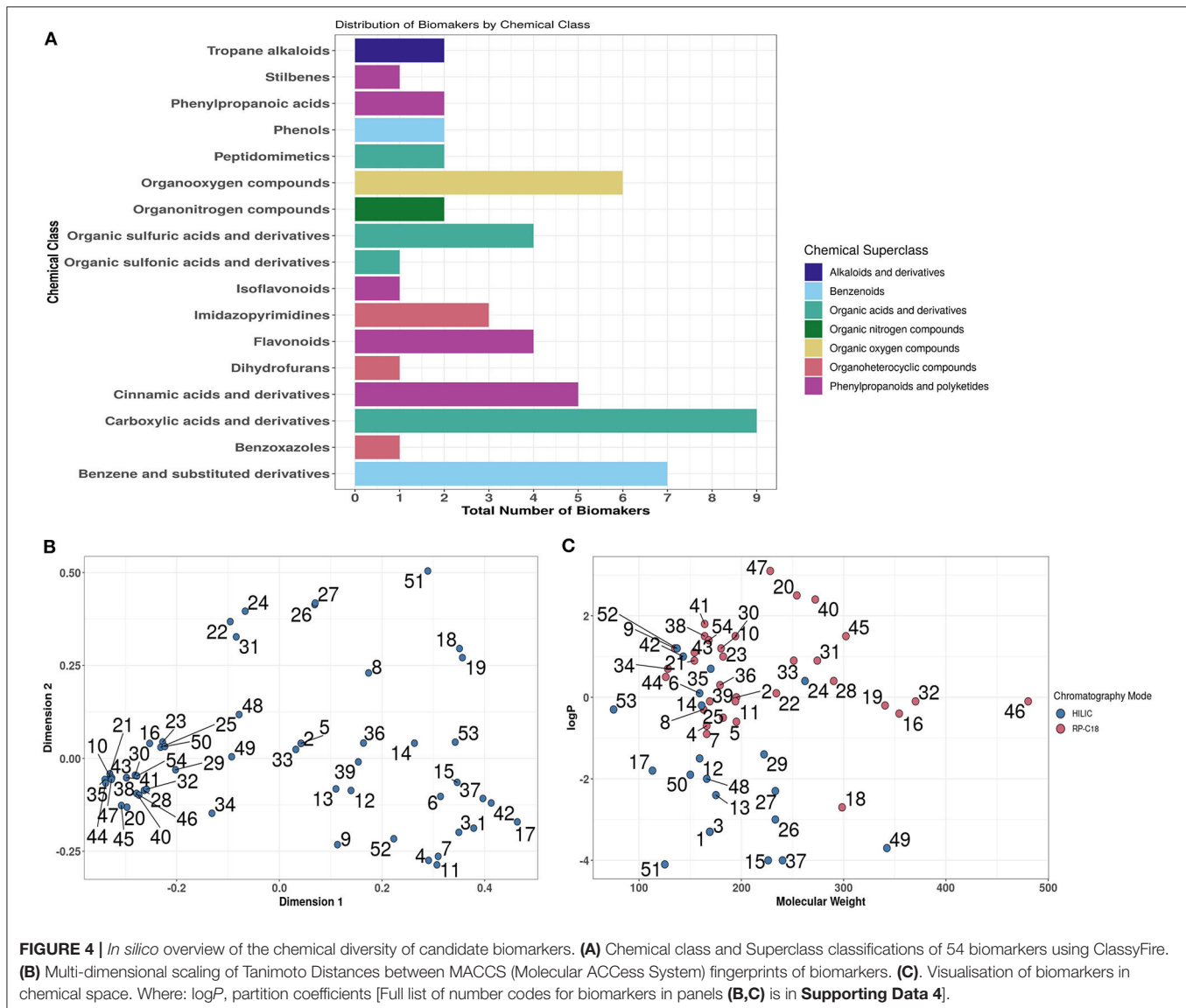
Dietary component	Putative urine biomarker	Potential use as a habitual dietary exposure biomarker?	Key reference in relation to habitual dietary exposure biomarker potential
Alcohol	Ethyl-beta-D-glucuronide	**	(78)
Wine	Resveratrol	*	(59)
Coffee	Chlorogenic acid	*	(79)
Coffee	Dihydrocaffeic acid [#]	*	(80)
Coffee	Ferulic acid-4-O-sulphate	*	(80)
Coffee	Feruloylglycine	**	(1)
Coffee	m-Coumaric acid [#]	*	(46)
Cocoa	3-Methyl-xanthine	**	(81)
Cocoa	7-Methyl-xanthine	**	(81)
Cocoa/Tea	Caffeic acid [#]	*	(80)
Coffee/Cocoa	Caffeine	*	(81)
Cocoa	Vanillic acid	**	(46)
Sweetener	Acesulphame-K	**	(2)
Sugary Foods and Drinks	Sucrose	**	(39)
Fruit and Vegetables	3-Hydroxyhippuric acid	**	(82)
Fruit and Vegetables	4-Hydroxyhippuric acid	**	(82)
Fruit and Vegetables	Hippuric acid	*	(83)
Citrus	4-Hydroxyproline-betaine	**	(30)
Banana	Dopamine-3-O-sulphate [#]	**	(84)
Banana	Dopamine-4-O-sulphate [#]	**	(84)
Strawberries/red berries	Furaneol	**	(2)
Citrus (grapefruit)	Naringenin	*	(85)
Grapes/wine/red berries	p-Coumaric acid	**	(86)
Citrus	Proline betaine	**	(30)
Apple	Rhamnitol	**	(22)
Grapes	Tartarate	**	(87)
Onion and tomato	Quercetin	*	(11)
Onion and tomato	Quercetin-3-O-b-D-glucuronide	**	(88)
Cruciferous Vegetables	D,L-Sulphoraphane L-cysteine	**	(89)
Cruciferous Vegetables	D,L-Sulphoraphane-N-acetyl-L-cysteine	**	(1)
Wholegrain/Rye	BOA (1,3-Benzoxazol-2-one)	*	(35)
Wholegrain	DHBA (3,5-Dihydroxybenzoic acid)	*	(90)
Wholegrain	DHBA-3-O-sulphate	*	(91)
Wholegrain	DHPPA (3-(3,5-Dihydroxyphenyl)-1-propanoic acid)	*	(90)
Wholegrain	DHPPA-3-sulphate	*	(91)
Meat (general)	1-Methyl histidine	*	(92)
Poultry/Fish	3-Methyl histidine	*	(27)
Meat (processed)	Carnitine	*	(93)
Red meat	Carnosine	*	(24)
Meat (general)	Creatinine	*	(25)
Chicken	L-Anserine	**	(1)
Meat (general)	Taurine	*	(25)
Fish/Shellfish	Trimethylamine-N-oxide	**	(1)
Potatoes	Calystegine A ₃	**	(48)
Potatoes	Calystegine B ₂ /B ₁	**	(48)
Soy products	Daidzein	**	(94)
Legumes	Pyrogallol	**	(2)
Legumes	Trigonelline	**	(2)

(Continued)

TABLE 2 | Continued

Dietary component	Putative urine biomarker	Potential use as a habitual dietary exposure biomarker?	Key reference in relation to habitual dietary exposure biomarker potential
Strongly Heated Foods	N-(2-Furoyl)glycine	**	(2)
Polyphenol rich foods	Epicatechin(-)	*	(79)
Polyphenol rich foods	Ferulic acid	*	(95)
Polyphenol/Anthocyanin rich foods	Ferulic acid-4-O-b-D-glucuronide	*	(96)
Fruit/Grapes/Tea/wine	Gallic acid	*	(86)
Anthocyanin rich foods	Protocatechuic acid	**	(97)

Normally the conjugated forms detected; the impact/use of the selected metabolite as a potential biomarker of habitual dietary exposure where *Possible and **Likely.



RSD data for biomarkers used to monitor exposure to six example foods/food groups are illustrated in **Figure 5** and in all cases, reproducibility gradually worsened as biomarker concentration

dropped. Median concentrations of the same biomarkers were measured in FMV urine taken from 95 free-living participants from **Study 1**. For each metabolite the median concentration was

substantially greater than the level at which RSD approached 20% (see dotted boxes in **Figure 5**) and usually an order of magnitude greater than the LoQ (see **Supporting Data 4**).

Demonstration of Biomarker Panel Utility to Examine Eating Behaviour in the MAIN Study

The utility of the biomarker panel to characterise eating habits within populations was explored by measuring the concentrations of 54 BFIs in FMV urine samples obtained on days following consumption of three distinctive meal plans (see text box in **Figure 6**) from 15 individuals in the Newcastle MAIN food intervention study (**Study 2**). The data were subjected to PCA which showed distinctive clustering of urine samples by Menu Day (colour coded) in relation to the zero position in PC1 and PC2 (indicated by dotted red grid lines in the scores plot shown in **Figure 6A**). Menu Day 1 and Menu Day 2 samples separated strongly in PC1, whereas Menu Day 3 samples clustered away from samples representative of the other 2 Menu Days in the PC2 dimension. Biomarkers that are strongly explanatory of differences in the composition of urines collected the morning after individual Menu Days are shown in **Figure 6B**. Examination of the 4 sectors delineated by the zero grid lines of the loadings plot revealed a strong association between specific biomarkers and particular foods consumed on each menu day. For example, TMAO (60) was strongly associated with cod fish fingers consumed on Menu Day 2, 3-Methyl histidine (3) was linked to chicken consumption on Menu Day 1 and carnosine (15) was indicative of exposure to a 100% beef burger on Menu Day 3. 1- and 3-Methyl-xanthine (4 and 7) and Epicatechin(-) (28; a marker for general polyphenol-rich foods) detected exposure to cocoa products on Menu Day 1, whilst Acesulphame-K (8) was associated with exposure to a diet soft drink on the same day. Sulphoraphane derivatives (18 and 19; D,L-Sulphoraphane L-cysteine and D,L-Sulphoraphane-N-acetyl-L-cysteine) were highly explanatory of exposure to coleslaw (containing cabbage) on Menu Day 3, whilst trigonelline (53) and N-(2-Furoyl)glycine (39; a strongly heated food marker) reflected exposure to coffee on the same day. Tartrate (51) and the calystegines (12 and 13; A₃ and B₂/B₁) contributed strongly to the clustering of urine samples from Menu Days 2 and 3 away from Menu Day 1 samples when grape products and potato products were not consumed. The example box plots (colour coded by Menu day) in **Figure 6C** demonstrate a clear increase in the concentration in urine of the selected biomarkers the day after the consumption of a specific food.

An important feature of any biomarker strategy designed to monitor habitual diet in both individuals and populations is the ability to add in new biomarkers as they are discovered and validated. RF can be used to assess the stringency of sample classification based on modelling output measures such as accuracy, AUC, and margins (76). **Figure 7** shows a MDS of proximity scores extracted from a RF classification model of the same 15 individuals consuming three unique menus, and panels of 38 dietary biomarkers used in 2018 (**Figure 7A**) and extended to 54 biomarkers in 2020 (**Figure 7B**). In both models, sample

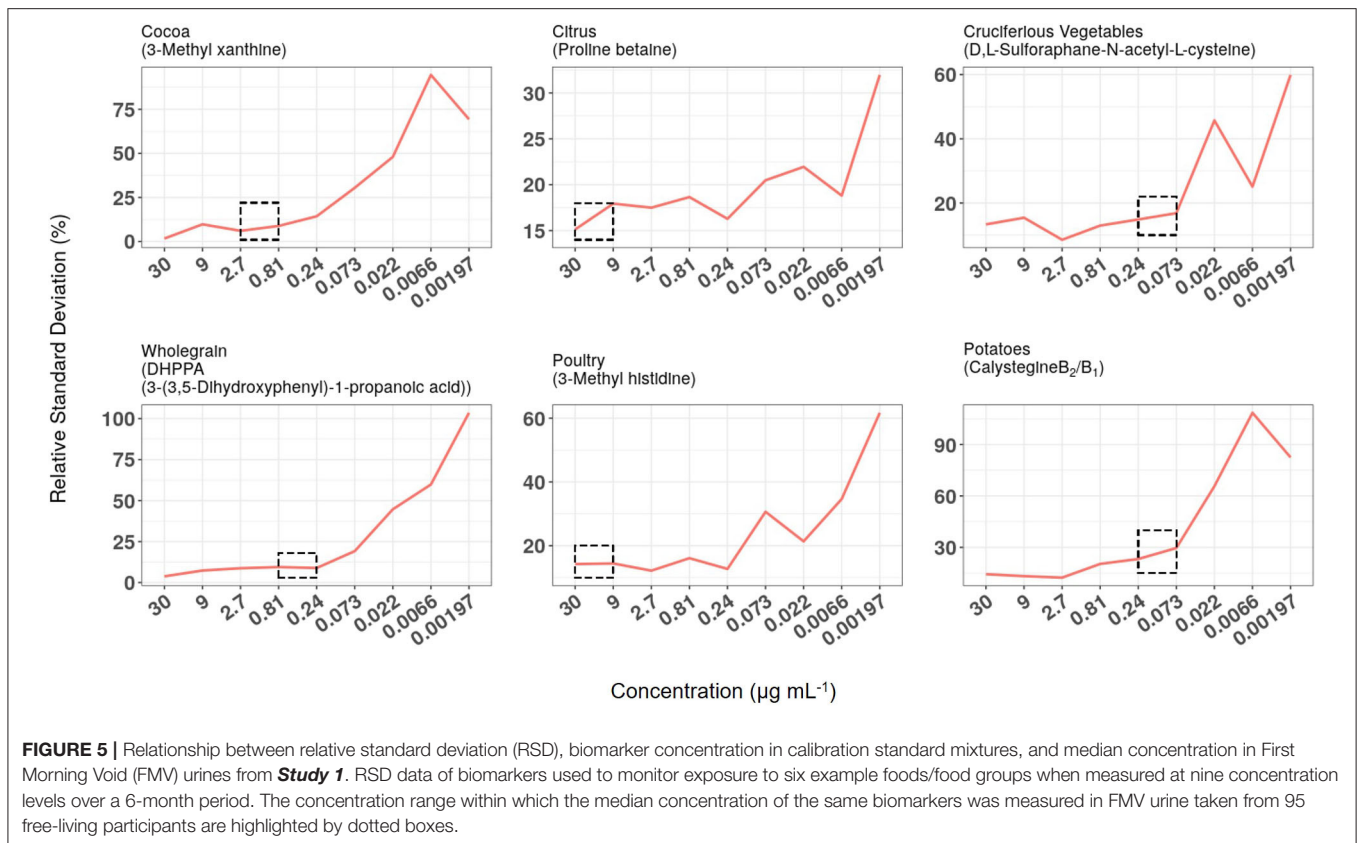
clustering by Menu Day is very similar and modelling output measures are still excellent, despite the challenge of measuring many more biomarkers in each MRM experiment in the more complex panel utilising 54 biomarkers.

DISCUSSION

More than a decade of intensive research to discover putative BFIs has yielded a wealth of information highlighting specific metabolites that appear in urine following consumption of individual foods/food groups. A large number of such metabolites have great potential as biomarker leads for particular foods/food groups when validated in isolation (49). However, their deployment in any cost-effective strategy aimed at comprehensive monitoring of habitual diet imposes a substantial number of further challenges that require both definition and investigation (1, 22). Preliminary investigation of the performance of any specific biomarker in the context of a complex biomarker panel requires urine samples from complex food intervention studies designed specifically to emulate habitual eating patterns. The MAIN Study at Newcastle was designed with this specific objective in mind (1, 2, 22). By validating biomarkers in urine samples from studies researching eating behaviour in free-living populations (62, 67) it is anticipated that BFI technology will mature rapidly over the next few years.

It is particularly important to consider carefully urine sampling approaches when deploying BFI technology to help monitor diet and to adopt a methodology that is appropriate for the study objectives (see **Table 1**). Twenty-four hour urine samples, which include the FMV after the study day, provide an ideal type of sample to assess food intake on a single day with the caveat that their collection imposes considerable burden on study participants. To assess habitual diet using BFI data would demand the collection 24 h urine samples on multiple days which can have a substantial influence on the acceptability of the study requirements and compliance by participants, as well as impacting greatly on study logistics and overall costs (1, 62, 67). Consequently, spot urine samples are becoming the urine samples of choice for studying BFIs of a single food/food group when compared with 24 h collections [e.g., (89, 98)] because their collection has little impact on normal daily activities of study participants.

Nutriketic studies of potential BFIs have shown that diet-derived metabolites from individual foods reach peak levels in urine at different times post-prandially (49). Thus, choosing an appropriately-timed spot urine sample is clearly problematical when considering the effective deployment of a comprehensive biomarker panel covering the whole of diet. Our recent studies have shown that spot urine samples are generally adequate substitutes for 24 h urine samples for measurement of BFIs (22, 62); particularly post-evening meal (i.e., just before bed time) and FMV urines were collected with a high degree of success (22). In the present paper, we describe a strategy to select urinary biomarkers for inclusion in a comprehensive and extendable panel to monitor habitual dietary exposure that focuses on the

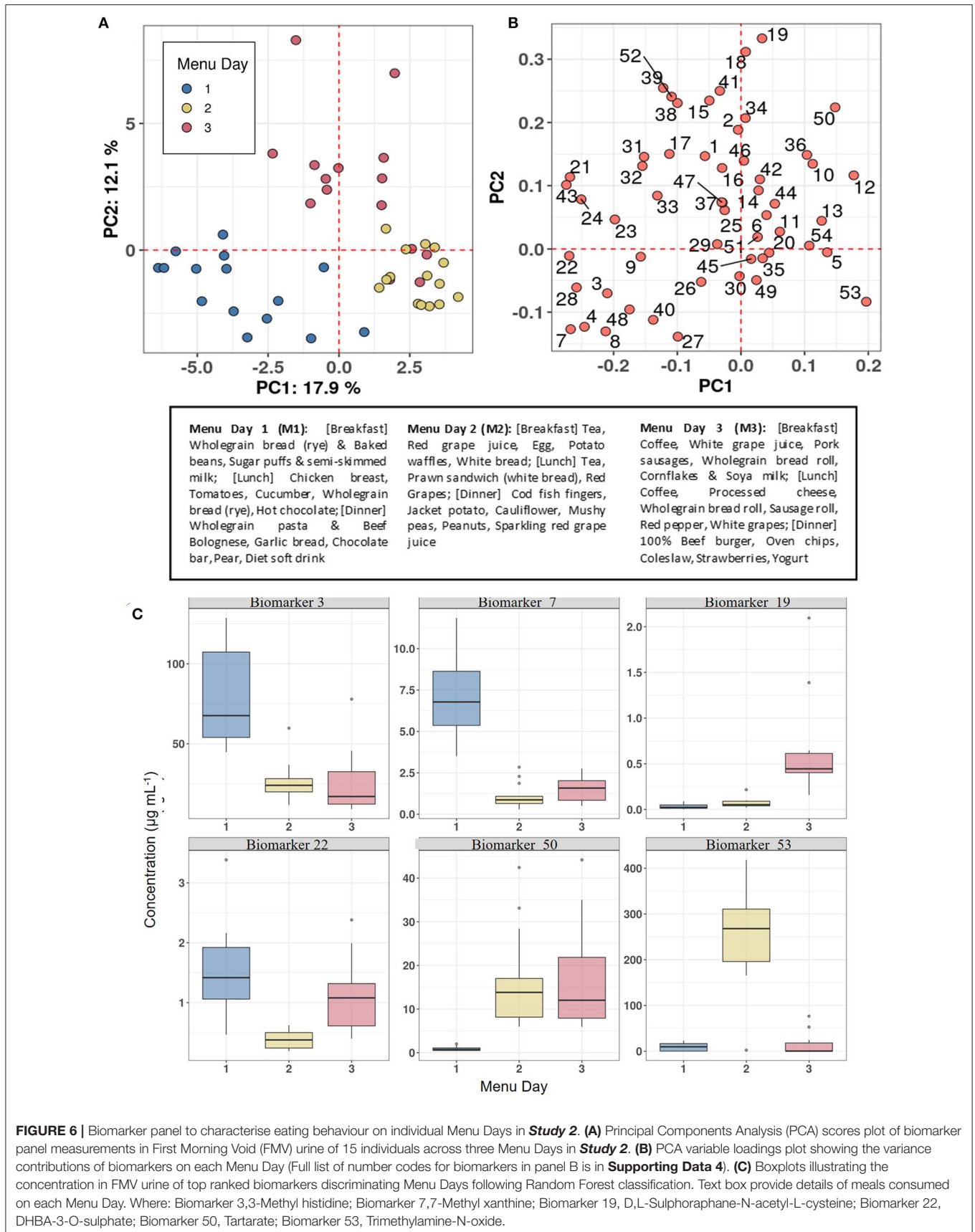


use of FMV urine samples. From an analytical perspective, sampling FMV urines after a substantial overnight sleep period allows sufficient time to elapse for any gut microbiome and liver P450 bio-transformations of targeted metabolites to be completed, thus extending the availability of characteristic biomarkers and increasing their concentrations in the collected urine. It has been shown previously that the distributions of biomarker concentrations are comparable between post-dinner spot samples, overnight cumulative samples, and 24 h urines (62). We demonstrated recently (53) that more than 50 different potential BFIs were detectable in FMV urine the day after the consumption of targeted foods. In the present study, we show that for many, but not all, BFIs there is a relatively linear correlation between concentration in 24 h and FMV urine and suggest that only those with an R^2 approaching 0.2 may be suitable for accurate quantification when using a comprehensive biomarker panel to monitor habitual diet. Although this is clearly a limitation for accurate quantification of food intake it is very likely that the presence in FMV urine of BFIs with lower correlation coefficients will still provide a useful qualitative indication of recent exposure to their target foods.

Our recent collaborations have highlighted the importance of understanding metabolic biotypes (metabotypes) in populations that may impact on nutritional status (41, 45, 99). As many dietary exposure biomarkers are derived from food chemicals that are metabolised and/or biotransformed before excretion, it is possible that chemical “signatures” reflective of common

metabotype groupings in any population can be visualised using a biomarker panel. Differential metabolism of any particular BFI by metabotype sub-groups in any population would provide an additional limitation on its utility for quantitative assessment of dietary intake. The hydration levels of study participants can vary considerably and has to be adjusted for in any BFI deployment strategy. The use of 24 h urine samples for biomarker quantification demands the accurate measurement of the total volume of urine produced during any 24 h period and then the concentration and extraction of a specific aliquot before analysis in order to calculate the overall daily excretion rate.

Logistics, the analytical and computational skills required, and costs will also impact on the wider acceptance and adoption of dietary exposure biomarker technology by the nutrition research community. In the methodology we describe, urine processing is limited to a simple dilution with ultra-pure water as QQQ instruments operating in MRM mode are extremely sensitive and thus there is a need to collect only small volumes of urine for analysis (e.g., 0.5–3 ml). With this objective in mind, we have shown that spot samples can be collected in the home with high collection compliance using vacuum tube technology (67). Importantly, urine samples collected by this method are compositionally stable at room temperature for several days without preservatives (67). This feature of vacuum tube collection methodology allows transport by domestic mail without dry ice offering the opportunity to scale up dietary exposure studies in community settings. A commercial



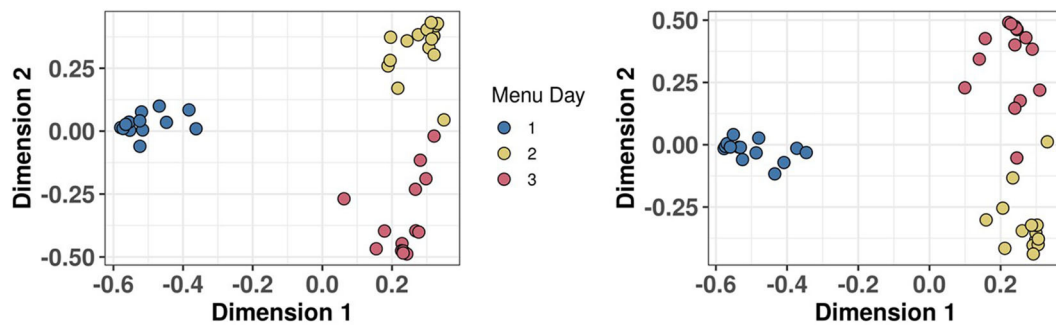


FIGURE 7 | Demonstration that biomarker panel can be extended without any loss of classification power. Multi-dimensional scaling (MDS) of Random Forest proximity extracted from a classification model of 15 individuals consuming three unique menus in **Study 2**. **(A)** MDS using a panel of 38 biomarkers of food intake (BFIs) which was extended to 54 BFIs and **(B)** 2 years later.

product for spot urine collection (**Supporting data 6**) is now on the market (<https://www.co-vertec.co.uk/>) and is currently under evaluation in several clinical trials interested in monitoring vulnerable populations in community environments to study malnutrition (<https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/stream-feasibility-study/>), impact of homelessness on diet (100) and evaluating the eating behaviours in pre-diabetic individuals (https://waru.org.uk/cms/waru_news/targeting-pre-diabetes-through-primary-care/).

Targeted profiling of urine using high resolution hybrid quadrupole/ion trap technology, coupled with RP C₁₈ UHPLC, can capture information about the relative concentrations of a substantial number of metabolites in a sample when combined with urine concentration by solid phase extraction methodology [e.g., (101)]. However, sample processing can add significant time and cost to any analytical process and in our experience differential metabolite recovery from ion exchange cartridges can add a significant degree of uncertainty and variance in metabolite measurement. In the present study, we have described a fully quantitative approach using QQQ-MS/MS to measure biomarker abundance. This methodology uses complex mixtures of chemicals standards for quantitation and utilises two HPLC columns solutions to provide optimal resolution of a structurally diverse range of chemicals using short chromatography runs.

As outlined in **Table 1** the utility of any biomarker panel will depend on the study objectives. The biomarker panel described in the present study was optimised specifically to investigate eating behaviour in free-living populations and was targeted towards frequently consumed foods of high public health importance in the UK (2). One limitation of the present study is that only relatively small populations have been used in these initial validation studies and in future BFI technology will need to be tested rigorously in multiple larger populations. We have shown that the biomarker panel can be extended incrementally as new biomarker leads are evaluated and current evidence suggests that it should be straightforward to adapt our strategy to develop biomarker panels that provide comprehensive coverage of foods consumed frequently in other

populations. Combined with existing bespoke software for data extraction, it is expected that the development of high throughput, automated biomarker measurement procedures to assess dietary intake is within scope in the near future. In addition, the routine generation of quantitative BFI data will offer further opportunities to develop novel “healthy eating indices” to summarise and “score” eating habits for use in personalised nutrition applications (26, 45). In conclusion, we believe that the integration of information from BFI technology and dietary self-reporting tools, combined with a deeper understanding of nutritional metabolic biotypes in populations, will help to provide more robust understanding of the complex interactions between dietary behaviour and human health.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Study 1 involving human participants were reviewed and approved by the National Commission for Data Protection, the Ethical Committee of the Institute of Public Health of the University of Porto and from the Ethical Commissions of each one of the Regional Administrations of Health. All participants gave written informed consent, and the study was carried out in accordance with the Declaration of Helsinki. Study 2 was reviewed and approved by East Midlands—Nottingham 1 National Research Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MB developed quantification methods, supervised MS support staff, and wrote the manuscript. TW developed quantification methods and wrote the manuscript. AL researched literature

and wrote the manuscript. NW undertook volunteer recruitment in Newcastle, coordinated volunteer CRF visits, and supervised CRF support staff. DT and AG undertook volunteer recruitment in Portugal and coordinated volunteer visits and supervised support staff. LL and HP provided QQQ technical support and data generation. JM coordinated project and supervised research in Newcastle University. JD coordinated project, supervised research in Aberystwyth, designed Figures, and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the Medical Research Council (MRC), grant number MR/J010308/1.

ACKNOWLEDGMENTS

The authors thank the participants in Newcastle and Portugal for their commitment in the food intervention and dietary survey studies, respectively. The researchers acknowledge all

the institutions and persons that integrated the activity of the IAN-AF (2015-2016) Consortium at all phases of the Survey.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnut.2020.602515/full#supplementary-material>

Supporting Data 1 | Putative urinary dietary exposure biomarkers from a literature search between the years of 2010–2020.

Supporting Data 2 | Foods consumed frequently by the UK population reported to have urine biomarkers and foods used in the MAIN study menu design.

Supporting Data 3 | Selection of biomarkers for panel development.

Supporting Data 4 | Technical detail of Triple Quadrupole targeted quantification of dietary intake biomarkers including: ionisation mode; chromatography column; retention time selected reaction monitoring (SRM) parameters of collision energy and product ion mass; quantification ion product; limitation of measurement, and concentrations in FMV urine.

Supporting Data 5 | Quantile-Quantile (QQ) plots of biomarker concentrations ($n = 52$), where **(A)** First Morning Void (FMV); **(B)** 24 h urine collection.

Supporting Data 6 | Urine collection kit characteristics and supplier.

REFERENCES

- Lloyd AJ, Willis ND, Wilson T, Zubair H, Xie L, Chambers E, et al. Developing a food exposure and urine sampling strategy for dietary exposure biomarker validation in free-living individuals. *Mol Nutr Food Res*. (2019) 63:1900062. doi: 10.1002/mnfr.201900062
- Willis ND, Lloyd AJ, Xie L, Stiegler M, Tailliant K, Garcia-Perez I, et al. Design and characterization of a randomized food intervention that mimics exposure to a typical UK diet to provide urine samples for identification and validation of metabolite biomarkers of food intake. *Front Nutr*. (2020) 7:51010. doi: 10.3389/fnut.2020.561010
- Penn L, Boeing H, Boushey CJ, Dragsted LO, Kaput J, Scalbert A, et al. Assessment of dietary intake: NuGO symposium report. *Genes Nutr*. (2010) 5:205–13. doi: 10.1007/s12263-010-0175-9
- Bingham SA, Gill C, Welch A, Day K, Cassidy A, Khaw KT, et al. Comparison of dietary assessment methods in nutritional epidemiology: weighed records v. 24 h recalls, food-frequency questionnaires and estimated-diet records. *Br J Nutr*. (1994) 72:619–43. doi: 10.1079/BJN19940064
- Lennox A, Bluck L, Page P, Pell D, Cole D, Ziauddeen N, et al. *Appendix X Misreporting in the National Diet and Nutrition Survey Rolling Programme (NDNS RP): Summary of Results and Their Interpretation*. National Diet and Nutrition Survey. Results from Years 1-4 (combined) of the Rolling Programme (2008/2009–2011/12) (2014).
- Poslusna K, Ruprich J, de Vries JH, Jakubikova M, van't Veer P. Misreporting of energy and micronutrient intake estimated by food records and 24 hour recalls, control and adjustment methods in practice. *Br J Nutr*. (2009) 101:S73–85. doi: 10.1017/S000711451004237
- Jessri M, Lou YW, L'Abbe MR. Evaluation of different methods to handle misreporting in obesity research: evidence from the Canadian national nutrition survey. *Br J Nutr*. (2016) 115:147–59. doi: 10.1017/S0007114515004237
- Münger LH, Garcia-Aloy M, Vázquez-Fresno R, Gille D, Rosana ARR, Passerini A, et al. Biomarker of food intake for assessing the consumption of dairy and egg products. *Genes Nutr*. (2018) 13:26. doi: 10.1186/s12263-018-0615-5
- Praticó G, Gao Q, Manach C, Dragsted LO. Biomarkers of food intake for Allium vegetables. *Genes Nutr*. (2018) 13:34. doi: 10.1186/s12263-018-0624-4
- Rothwell JA, Madrid-Gambin F, Garcia-Aloy M, Andres-Lacueva C, Logue C, Gallagher AM, et al. Biomarkers of intake for coffee, tea, and sweetened beverages. *Genes Nutr*. (2018) 13:15. doi: 10.1186/s12263-018-0607-5
- Sri Harsha PSC, Wahab RA, Garcia-Aloy M, Madrid-Gambin F, Estruel-Amades S, Watzl B, et al. Biomarkers of legume intake in human intervention and observational studies: a systematic review. *Genes Nutr*. (2018) 13:25. doi: 10.1186/s12263-018-0616-4
- Ulaszewska M, Vázquez-Manjarrez N, Garcia-Aloy M, Llorach R, Mattivi F, Dragsted LO, et al. Food intake biomarkers for apple, pear, and stone fruit. *Genes Nutr*. (2018) 13:29. doi: 10.1186/s12263-018-0620-8
- Cuparencu C, Praticó G, Hemeryck LY, Sri Harsha PSC, Noerman S, Rombouts C, et al. Biomarkers of meat and seafood intake: an extensive literature review. *Genes Nutr*. (2019) 14:35. doi: 10.1186/s12263-019-0656-4
- Garcia-Aloy M, Hulshof PJM, Estruel-Amades S, Osté MCJ, Lankinen M, Geleijnse JM, et al. Biomarkers of food intake for nuts and vegetable oils: an extensive literature search. *Genes Nutr*. (2019) 14:7. doi: 10.1186/s12263-019-0628-8
- Landberg R, Hanhineva K, Tuohy K, Garcia-Aloy M, Biskup I, Llorach R, et al. Biomarkers of cereal food intake. *Genes Nutr*. (2019) 14:28. doi: 10.1186/s12263-019-0651-9
- Maruvada P, Lampe JW, Wishart DS, Barupal D, Chester DN, Dodd D, et al. Perspective: dietary biomarkers of intake and exposure-exploration with omics approaches. *Adv Nutr*. (2019) 11:200–15. doi: 10.1093/advances/nmz075
- Zhou X, Gao Q, Praticó G, Chen J, Dragsted LO. Biomarkers of tuber intake. *Genes Nutr*. (2019) 14:9. doi: 10.1186/s12263-019-0631-0
- Ulaszewska M, Garcia-Aloy M, Vázquez-Manjarrez N, Soria-Flórido MT, Llorach R, Mattivi F, et al. Food intake biomarkers for berries and grapes. *Genes Nutr*. (2020) 15:17. doi: 10.1186/s12263-020-00675-z
- Brouwer-Brolsma EM, Brennan L, Drevon CA, van Kranen H, Manach C, Dragsted LO, et al. Combining traditional dietary assessment methods with novel metabolomics techniques: present efforts by the Food Biomarker Alliance. *Proc Nutr Soc*. (2017) 76:619–27. doi: 10.1017/S0029665117003949
- Lloyd AJ, Beckmann M, Haldar S, Seal C, Brandt K, Draper J. Data-driven strategy for the discovery of potential urinary biomarkers of habitual dietary exposure. *Am J Clin Nutr*. (2013) 97:377–89. doi: 10.3945/ajcn.112.048033
- Department of Health. *National Diet and Nutrition Survey: Headline Results from Years 1, 2 and 3 (Combined) of the Rolling Programme 2008/09 - 2010/11*. (2012) (Retrieved June, 2017).
- Lloyd AJ, Willis ND, Wilson T, Zubair H, Chambers E, Garcia-Perez I, et al. Addressing the pitfalls when designing intervention studies to discover and validate biomarkers of habitual dietary intake. *Metabolomics*. (2019) 15:72. doi: 10.1007/s11306-019-1532-3

23. Favé G, Beckmann ME, Draper JH, Mathers JC. Measurement of dietary exposure: a challenging problem which may be overcome thanks to metabolomics? *Genes Nutr.* (2009) 4:135–41. doi: 10.1007/s12263-009-0120-y
24. Yeum KJ, Orioli M, Regazzoni L, Carini M, Rasmussen H, Russell RM, et al. Profiling histidine dipeptides in plasma and urine after ingesting beef, chicken or chicken broth in humans. *Amino Acids.* (2010) 38:847–58. doi: 10.1007/s00726-009-0291-2
25. Cross AJ, Major MJ, Sinha R. Urinary biomarkers of meat consumption. *Cancer Epidemiol Biomark Prev.* (2011) 20:1107–11. doi: 10.1158/1055-9965.EPI-11-0048
26. Playdon MC, Moore SC, Derkach A, Reedy J, Subar AF, Sampson JN, et al. Identifying biomarkers of dietary patterns by using metabolomics. *Am J Clin Nutr.* (2016) 105:450–65. doi: 10.3945/ajcn.116.144501
27. Cheung W, Keski-Rahkonen P, Assi N, Ferrari P, Freisling H, Rinaldi S, et al. A metabolomic study of biomarkers of meat and fish intake. *Am J Clin Nutr.* (2017) 105:600–8. doi: 10.3945/ajcn.116.146639
28. Cuparencu C, Rinnan Å, Dragsted LO. Combined markers to assess meat intake-human metabolomic studies of discovery and validation. *Mol Nutr Food Res.* (2019) 63:1900106. doi: 10.1002/mnfr.201900106
29. Heinzmann SS, Brown IJ, Chan Q, Bictash M, Dumas M-E, Kochhar S, et al. Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. *Am J Clin Nutr.* (2010) 92:436–43. doi: 10.3945/ajcn.2010.29672
30. Lloyd AJ, Beckmann M, Fave G, Mathers CJ, Draper J. Proline betaine and its biotransformation products in fasting urine samples are potential biomarkers of habitual citrus fruit consumption. *Br J Nutr.* (2011) 106:812–24. doi: 10.1017/S0007114511001164
31. Edmands WMB, Beckonert OP, Stella C, Campbell A, Lake BG, Lindon JC, et al. Identification of human urinary biomarkers of cruciferous vegetable consumption by metabolomic profiling. *J Proteome Res.* (2011) 10:4513–21. doi: 10.1021/pr200326k
32. Lloyd AJ, Fave G, Beckmann M, Lin W, Tailliant K, Xie L, et al. Use of mass spectrometry fingerprinting to identify urinary metabolites after consumption of specific foods. *Am J Clin Nutr.* (2011) 94:981–91. doi: 10.3945/ajcn.111.017921
33. González-Barrio R, Borges G, Mullen, W, Crozier A. Bioavailability of anthocyanins and ellagitannins following consumption of raspberries by healthy humans and subjects with an ileostomy. *J Agric Food Chem.* (2010) 58:3933–39. doi: 10.1021/jf100315d
34. Cuparencu CS, Andersen M-BS, Gürdeniz G, Schou SS, Mortensen MW, Raben A, et al. Identification of urinary biomarkers after consumption of sea buckthorn and strawberry, by untargeted LC-MS metabolomics: a meal study in adult men. *Metabolomics.* (2016) 12:31. doi: 10.1007/s11306-015-0934-0
35. Beckmann M, Lloyd AJ, Haldar S, Seal C, Brandt K, Draper J. Hydroxylated phenylacetamides derived from bioactive benzoxazinoids are bioavailable in humans after habitual consumption of whole grain sourdough rye bread. *Mol Nutr Food Res.* (2013) 57:1859–73. doi: 10.1002/mnfr.201200777
36. Bondia-Pons I, Barri T, Hanhineva K, Juntunen K, Dragsted LO, Mykkänen H, et al. UPLC-QTOF/MS metabolic profiling unveils urinary changes in humans after a whole grain rye versus refined wheat bread intervention. *Mol Nutr Food Res.* (2013) 57:412–22. doi: 10.1002/mnfr.201200571
37. Garcia-Aloy M, Llorach R, Urpi-Sarda M, Tulipani S, Salas-Salvadó J, Martínez-González MA, et al. Nutrimetabolomics fingerprinting to identify biomarkers of bread exposure in a free-living population from the PREDIMED study cohort. *Metabolomics.* (2015) 11:155–65. doi: 10.1007/s11306-014-0682-6
38. Andersen M-BS, Rinnan Å, Manach C, Poulsen SK, Pujos-Guillot E, Larsen TM, et al. Untargeted metabolomics as a screening tool for estimating compliance to a dietary pattern. *J Proteome Res.* (2014) 13:1405–18. doi: 10.1021/pr400964s
39. Beckmann M, Joosen AM, Clarke MM, Mugridge O, Frost G, Engel B, et al. Changes in the human plasma and urinary metabolome associated with acute dietary exposure to sucrose and the identification of potential biomarkers of sucrose intake. *Mol Nutr Food Res.* (2016) 60:444–57. doi: 10.1002/mnfr.201500495
40. Logue C, Dowe LRC, Strain JJ, Verhagen H, McClean S, Gallagher AM. Application of liquid chromatography-tandem mass spectrometry to determine urinary concentrations of five commonly used low-calorie sweeteners: a novel biomarker approach for assessing recent intakes? *J Agric Food Chem.* (2017) 65:4516–25. doi: 10.1021/acs.jafc.7b00404
41. Posma JM, Garcia-Perez I, Heaton JC, Burdisso P, Mathers JC, Draper J, et al. Integrated analytical and statistical two-dimensional spectroscopy strategy for metabolite identification: application to dietary biomarkers. *Anal Chem.* (2017) 89:3300–9. doi: 10.1021/acs.analchem.6b03324
42. Garcia-Aloy M, Ulaszewska M, Franceschi P, Estruel-Amades S, Weinert CH, Tor-Roca A, et al. Discovery of intake biomarkers of lentils, chickpeas, and white beans by untargeted LC-MS metabolomics in serum and urine. *Mol Nutr Food Res.* (2020) 64:e1901137. doi: 10.1002/mnfr.201901137
43. Stalmach A, Edwards CA, Wightman, DJ, Crozier A. Identification of (poly)phenolic compounds in concord grape juice and their metabolites in human plasma and urine after juice consumption. *J Agric Food Chem.* (2011) 59:9512–22. doi: 10.1021/jf2015039
44. Khymentis O, Andres-Lacueva C, Urpi-Sarda M, Vazquez-Fresno R, Mart MM, Reglero G, et al. Metabolic fingerprint after acute and under sustained consumption of a functional beverage based on grape skin extract in healthy human subjects. *Food Funct.* (2015) 6:1288–98. doi: 10.1039/C4FO00684D
45. Garcia-Perez I, Posma JM, Gibson R, Chambers ES, Hansen TH, Vestergaard H, et al. Objective assessment of dietary patterns by use of metabolic phenotyping: a randomised, controlled, crossover trial. *Lancet Diabetes Endocrinol.* (2017) 5:184–95. doi: 10.1016/S2213-8587(16)30419-3
46. Edmands WM, Ferrari P, Rothwell JA, Rinaldi S, Slimani N, Barupal DK, et al. Polyphenol metabolome in human urine and its association with intake of polyphenol-rich foods across European countries. *Am J Clin Nutr.* (2015) 102:905–13. doi: 10.3945/ajcn.114.101881
47. Ulaszewska MM, Koutsos A, Trošt K, Stanstrup J, Garcia-Aloy M, Scholz M, et al. Two apples a day modulate human: microbiome co-metabolic processing of polyphenols, tyrosine and tryptophan. *Eur J Nutr.* (2020). doi: 10.1007/s00394-020-02201-8. [Epub ahead of print].
48. Beckmann M, Lloyd AJ, Wilson T, Torres D, Goios A, Willis N. Calystegines are potential urine biomarkers for dietary exposure to potato products. *Mol Nutr Food Res.* (2020) 64:2000515. doi: 10.1002/mnfr.202000515
49. Dragsted LO, Gao Q, Scalbert A, Vergères G, Kolehmainen M, Manach C, et al. Validation of biomarkers of food intake-critical assessment of candidate biomarkers. *Genes Nutr.* (2018) 13:14. doi: 10.1186/s12263-018-0603-9
50. Kristensen M, Engelsens SB, Dragsted LO. LC-MS metabolomics top-down approach reveals new exposure and effect biomarkers of apple and apple-pectin intake. *Metabolomics.* (2012) 8:64–73. doi: 10.1007/s11306-011-0282-7
51. Praticò G, Gao Q, Scalbert A, Vergères G, Kolehmainen M, Manach C, et al. Guidelines for Biomarker of Food Intake Reviews (BFIRev): how to conduct an extensive literature search for biomarker of food intake discovery. *Genes Nutr.* (2018) 13:3. doi: 10.1186/s12263-018-0592-8
52. Garcia-Aloy M, Rabassa M, Casas-Agustench P, Hidalgo-Liberona N, Llorach R, Andres-Lacueva C. Novel strategies for improving dietary exposure assessment: multiple-data fusion is a more accurate measure than the traditional single-biomarker approach. *Trends Food Sci Technol.* (2017) 69:220–9. doi: 10.1016/j.tifs.2017.04.013
53. Beckmann M, Wilson T, Zubair H, Lloyd AJ, Lyons L, Phillips H, et al. A standardized strategy for simultaneous quantification of urine metabolites to validate development of a biomarker panel allowing comprehensive assessment of dietary exposure. *Mol Nutr Food Res.* (2020) 64:2000517. doi: 10.1002/mnfr.202000517
54. Quifer-Rada P, Martínez-Huelamo M, Chiva-Blanch G, Jáuregui O, Estruch R, Lamuela-Raventós RM. Urinary isoxanthohumol is a specific and accurate biomarker of beer consumption. *J Nutr.* (2014) 144:484–8. doi: 10.3945/jn.113.185199
55. Vetrani C, Rivellese AA, Annuzzi G, Mattila I, Meudec E, Hyötyläinen T, et al. Phenolic metabolites as compliance biomarker for polyphenol intake in a randomized controlled human intervention. *Food Res. Int.* (2014) 63:233–8. doi: 10.1016/j.foodres.2014.01.018
56. Lang R, Lang T, Bader M, Beusch A, Schlagbauer V, Hofmann T. High-throughput quantitation of proline betaine in foods and suitability as a valid

- biomarker for citrus consumption. *J Agric Food Chem.* (2017) 65:1613–9. doi: 10.1021/acs.jafc.6b05824
57. O'Donovan CB, Walsh MC, Woolhead C, Forster H, Celis-Morales C, Fallaize R, et al. Metabotyping for the development of tailored dietary advice solutions in a European population: the Food4Me study. *Br J Nutr.* (2017) 118:561–9. doi: 10.1017/S0007114517002069
 58. Estruch R, Martínez-González MA, Corella D, Salas-Salvadó J, Ruiz-Gutiérrez V, Covas MI, et al. Effects of a Mediterranean-style diet on cardiovascular risk factors: a randomized trial. *Ann Intern Med.* (2006) 145:1–11. doi: 10.7326/0003-4819-145-1-200607040-00004
 59. Zamora-Ros R, Rothwell JA, Achaintre D, Ferrari P, Boutron-Ruault M-C, Mancini FR, et al. Evaluation of urinary resveratrol as a biomarker of dietary resveratrol intake in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Br J Nutr.* (2017) 117:1596–1602. doi: 10.1017/S0007114517001465
 60. Lopes C, Torres D, Oliveira A, Severo M, Guiomar S, Alarcão V, et al. National food, nutrition, and physical activity survey of the portuguese general population (2015–2016): protocol for design and development. *JMIR Res Protoc.* (2018) 7:e42. doi: 10.2196/resprot.8990
 61. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. European prospective investigation into cancer and nutrition (EPIC): study populations and data collection. *Public Health Nutr.* (2002) 5:1113–24. doi: 10.1079/PHN2002394
 62. Wilson T, Garcia-Perez I, Posma JM, Lloyd AJ, Chambers ES, Tailliant K, et al. Spot and cumulative urine samples are suitable replacements for 24-h urine collections for objective measures of dietary exposure in adults using metabolite biomarkers. *J Nutr.* (2019) 149:1692–700. doi: 10.1093/jn/nxz138
 63. Hong YH, Dublin N, Razack AH, Mohd AM, Husain R. Twenty-four hour and spot urine metabolic evaluations: correlations versus agreements. *Urology.* (2010) 75:1294–8. doi: 10.1016/j.urology.2009.08.061
 64. Bradman A, Kogut K, Eisen EA, Jewell NP, Quirós-Alcalá L, Castorina R, et al. Variability of organophosphorus pesticide metabolite levels in spot and 24-hr urine samples collected from young children during 1 week. *Environ Health Perspect.* (2013) 121:118–24. doi: 10.1289/ehp.1104808
 65. Wang Y-X, Feng W, Zeng Q, Sun Y, Wang P, You L, et al. Variability of metal levels in spot, first morning, and 24-hour urine samples over a 3-month period in healthy adult Chinese men. *Environ Health Perspect.* (2016) 124:468–76. doi: 10.1289/ehp.1409551
 66. Edmands WMB, Ferrari P, Scalbert A. Normalization to specific gravity prior to analysis improves information recovery from high resolution mass spectrometry metabolomic profiles of human urine. *Anal Chem.* (2014) 86:10925–31. doi: 10.1021/ac503190m
 67. Lloyd AJ, Wilson T, Willis ND, Lyons L, Phillips H, Janssen HG, et al. Developing community-based urine sampling methods to deploy biomarker technology for the assessment of dietary exposure. *Public Health Nutr.* (2020) 1–12. doi: 10.1017/S136898002000097X. [Epub ahead of print].
 68. Goios ACL, Severo M, Lloyd AJ, Magalhães VPL, Lopes C, Torres DPM. Validation of a new software eAT24 used to assess dietary intake in the adult Portuguese population. *Public Health Nutr.* (2020) 1–11. doi: 10.1017/S1368980020001044. [Epub ahead of print].
 69. Lopes C, Torres D, Oliveira A, Severo M, Guiomar S, Alarcão V, et al. National food, nutrition and physical activity survey of the portuguese general population. *EFSA Support Publ.* (2017) 14:37. doi: 10.2903/sp.efsa.2017.EN-1341
 70. Favé G, Beckmann M, Lloyd AJ, Zhou S, Harold G, Lin W, et al. Development and validation of a standardized protocol to monitor human dietary exposure by metabolite fingerprinting of urine samples. *Metabolomics.* (2011) 7:469–84. doi: 10.1007/s11306-011-0289-0
 71. Djoumbou Feunang Y, Eisner R, Knox C, Chepelev L, Hastings J, Owen G, et al. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J Cheminform.* (2016) 8:61. doi: 10.1186/s13321-016-0174-y
 72. Martens L, Chambers M, Sturm M, Kessner D, Levander F, Shofstahl J, et al. mzML—a community standard for mass spectrometry data. *Mol Cell Proteomics.* (2011) 10:R110.000133. doi: 10.1074/mcp.R110.000133
 73. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, et al. A Cross-platform Toolkit for Mass Spectrometry and Proteomics. *Nat Biotechnol.* (2012) 30:918–20. doi: 10.1038/nbt.2377
 74. R Core Team. *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing (2013).
 75. Liaw A, Wiener M. Classification and Regression by randomForest. *R News.* (2002) 2:18–22.
 76. Enot DP, Lin W, Beckmann M, Parker D, Overy PD, Draper J. Preprocessing, classification modeling and feature selection using flow injection electrospray mass spectrometry metabolite fingerprint data. *Nat Protoc.* (2008) 3:446–70. doi: 10.1038/nprot.2007.511
 77. Public Health England. *The Eatwell Guide.* (2016). Available online at: <https://www.gov.uk/government/publications/the-eatwell-guide> (Retrieved November, 2017).
 78. Dahl H, Stephanson N, Beck O, Helander A. Comparison of urinary excretion characteristics of ethanol and ethyl glucuronide. *J Anal Toxicol.* (2002) 26:201–4. doi: 10.1093/jat/26.4.201
 79. Ito H, Gonthier M-P, Manach C, Morand C, Mennen L, Rémésy C, et al. Polyphenol levels in human urine after intake of six different polyphenol-rich beverages. *Br J Nutr.* (2007) 94:500–9. doi: 10.1079/BJN20051522
 80. Stalmach A, Mullen W, Barron D, Uchida K, Yokota T, Cavin C, et al. Metabolite profiling of hydroxycinnamate derivatives in plasma and urine after the ingestion of coffee by humans: identification of biomarkers of coffee consumption. *Drug Metab Dispos.* (2009) 37:1749–58. doi: 10.1124/dmd.109.028019
 81. Martínez-López S, Sarriá B, Gómez-Juaristi M, Goya L, Mateos R, Bravo-Clemente L. Theobromine, caffeine, and theophylline metabolites in human plasma and urine after consumption of soluble cocoa products with different methylxanthine contents. *Food Res. Int.* (2014) 63:446–55. doi: 10.1016/j.foodres.2014.03.009
 82. Rechner AR, Kuhnle G, Hu H, Roedig-Penman A, van den Braak MH, Moore KP, et al. The metabolism of dietary polyphenols and the relevance to circulating levels of conjugated metabolites. *Free Rad Res.* (2002) 36:1229–41. doi: 10.1080/246-1071576021000016472
 83. 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: 1314–20. doi: 10.3945/jn.112.159319
 84. Vázquez-Manjarrez N, Weinert CH, Ulaszewska MM, Mack CI, Micheau P, Pétera M, et al. Discovery and validation of banana intake biomarkers using untargeted metabolomics in human intervention and cross-sectional studies. *J Nutr.* (2019) 149:1701–13. doi: 10.1093/jn/nxz125
 85. Erlund I, Meririnne E, Alftan G, Aro A. Plasma kinetics and urinary excretion of the flavanones naringenin and hesperetin in humans after ingestion of orange juice and grapefruit juice. *J Nutr.* (2001) 131:235–41. doi: 10.1093/jn/131.2.235
 86. Boto-Ordóñez M, Urpi-Sarda M, Queipo-Ortuño MI, Corella D, Tinahones FJ, Estruch R, et al. Microbial metabolomic fingerprinting in urine after regular dealcoholized red wine consumption in humans. *J Agric Food Chem.* (2013) 61:9166–75. doi: 10.1021/jf402394c
 87. Garcia-Perez I, Posma JM, Chambers ES, Nicholson JK, Mathers JC, Beckmann M, et al. An analytical pipeline for quantitative characterization of dietary intake: application to assess grape intake. *J Agric Food Chem.* (2016) 64:2423–31. doi: 10.1021/acs.jafc.5b05878
 88. Mullen W, Edwards CA, Crozier A. Absorption, excretion and metabolite profiling of methyl-, glucuronyl-, glucosyl- and sulpho-conjugates of quercetin in human plasma and urine after ingestion of onions. *Br J Nutr.* (2007) 96:107–16. doi: 10.1079/BJN20061809
 89. Haider J, Winkler S, Bub A, Rüfer CE, Pignitter M, Somoza V, et al. LC-MS/MS quantification of sulforaphane and indole-3-carbinol metabolites in human plasma and urine after dietary intake of selenium-fortified broccoli. *J Agric Food Chem.* (2011) 59:8047–57. doi: 10.1021/jf201501x
 90. Zhu Y, Shurlock KL, Chen X, Sang S. Identification and pharmacokinetics of novel alkylresorcinol metabolites in human urine, new candidate biomarkers for whole-grain wheat and rye intake. *J Nutr.* (2014) 144:114–22. doi: 10.3945/jn.113.184663
 91. Zhu Y, Wang P, Sha W, Sang S. Urinary biomarkers of whole grain wheat intake identified by non-targeted and targeted metabolomics approaches. *Sci Rep.* (2016) 6:36278. doi: 10.1038/srep36278
 92. Altorf-van der Kuil W, Brink EJ, Boetje M, Siebelink E, Bijlsma S, Engberink MF, et al. Identification of biomarkers for intake of protein from meat, dairy

- products and grains: a controlled dietary intervention study. *Br J Nutr.* (2013) 110:810–22. doi: 10.1017/S0007114512005788
93. Heinzmann SS, Merrifield CA, Rezzi S, Kochhar S, Lindon JC, Holmes E, et al. Stability and robustness of human metabolic phenotypes in response to sequential food challenges. *J Proteome Res.* (2012) 11:643–55. doi: 10.1021/pr2005764
94. King RA, Bursill DB. Plasma and urinary kinetics of the isoflavones daidzein and genistein after a single soy meal in humans. *Am J Clin Nutr.* (1998) 67:867–72. doi: 10.1093/ajcn/67.5.867
95. Gonthier M-P, Rios LY, Verny M-A, Révész C, Scalbert A. Novel liquid chromatography-electrospray ionization mass spectrometry method for the quantification in human urine of microbial aromatic acid metabolites derived from dietary polyphenols. *J Chromatogr B.* (2003) 789:247–55. doi: 10.1016/S1570-0232(03)00073-4
96. Agulló V, Villaño D, García-Viguera C, Domínguez-Perles R. Anthocyanin metabolites in human urine after the intake of new functional beverages. *Molecules.* (2020) 25:371. doi: 10.3390/molecules25020371
97. Nurmi T, Mursu J, Heinonen M, Nurmi A, Hiltunen R, Voutilainen S. Metabolism of berry anthocyanins to phenolic acids in humans. *J Agric Food Chem.* (2009) 57:2274–81. doi: 10.1021/jf8035116
98. Urbain, P, Bertz H. Monitoring for compliance with a ketogenic diet: what is the best time of day to test for urinary ketosis? *Nutr Metab.* (2016) 13:77. doi: 10.1186/s12986-016-0136-4
99. Garcia-Perez I, Posma JM, Chambers ES, Mathers JC, Draper J, Beckmann M, et al. Dietary metabotype modelling predicts individual responses to dietary interventions. *Nature Food.* (2020) 1:355–64. doi: 10.1038/s43016-020-0092-z
100. Fallaize R, Seale JV, Mortin C, Armstrong L, Lovegrove JA. Dietary intake, nutritional status and mental wellbeing of homeless adults in Reading, UK. *Br J Nutr.* (2017) 118:707–14. doi: 10.1017/S0007114517002495
101. González-Domínguez R, Urpi-Sarda M, Jáuregui O, Needs PW, Kroon AP, Andrés-Lacueva C. Quantitative dietary fingerprinting (QDF)-a novel tool for comprehensive dietary assessment based on urinary nutrimentabolomics. *J Agric Food Chem.* (2020) 68:1851–61. doi: 10.1021/acs.jafc.8b07023

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Beckmann, Wilson, Lloyd, Torres, Goios, Willis, Lyons, Phillips, Mathers and Draper. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.