



Conference on 'Translating nutrition: integrating research, practice and policy' Symposium 2: Intervention study design and personalised nutrition

Dietary exposure biomarker-lead discovery based on metabolomics analysis of urine samples

Manfred Beckmann¹, Amanda J. Lloyd¹, Sumanto Haldar², Gaëlle Favé³, Chris J. Seal²,
Kirsten Brandt², John C. Mathers³ and John Draper^{1*}

¹Institute of Biological Environmental and Rural Sciences, Aberystwyth University, Aberystwyth SY23 3DA, UK

²Human Nutrition Research Centre, School of Agriculture, Food and Rural Development, Newcastle University,
Newcastle upon Tyne NE1 7RU, UK

³Human Nutrition Research Centre, Institute for Ageing and Health, Campus for Ageing and Vitality, Newcastle University,
Newcastle upon Tyne NE4 5PL, UK

Although robust associations between dietary intake and population health are evident from conventional observational epidemiology, the outcomes of large-scale intervention studies testing the causality of those links have often proved inconclusive or have failed to demonstrate causality. This apparent conflict may be due to the well-recognised difficulty in measuring habitual food intake which may lead to confounding in observational epidemiology. Urine biomarkers indicative of exposure to specific foods offer information supplementary to the reliance on dietary intake self-assessment tools, such as FFQ, which are subject to individual bias. Biomarker discovery strategies using non-targeted metabolomics have been used recently to analyse urine from either short-term food intervention studies or from cohort studies in which participants consumed a freely-chosen diet. In the latter, the analysis of diet diary or FFQ information allowed classification of individuals in terms of the frequency of consumption of specific diet constituents. We review these approaches for biomarker discovery and illustrate both with particular reference to two studies carried out by the authors using approaches combining metabolite fingerprinting by MS with supervised multivariate data analysis. In both approaches, urine signals responsible for distinguishing between specific foods were identified and could be related to the chemical composition of the original foods. When using dietary data, both food distinctiveness and consumption frequency influenced whether differential dietary exposure could be discriminated adequately. We conclude that metabolomics methods for fingerprinting or profiling of overnight void urine, in particular, provide a robust strategy for dietary exposure biomarker-lead discovery.

Dietary exposure: Metabolite fingerprinting: FFQ: Multivariate data analysis: Urine biomarkers

The accurate measurement of dietary exposure, which is an essential component of much health-related research, offers a challenging prospect. Specifically, dietary data can be subject to participant bias and can depend heavily upon food composition tables for the estimation of intakes of energy, nutrients and other food constituents^(1–6). The chemical content of body fluids is a potentially rich source

of information about dietary exposure as many foods contain distinctive metabolites which give rise to further chemical diversity following food ingestion, absorption and metabolism^(7–10). However, to date, putative biochemical markers are available for only a relatively small number of specific foods and food components⁽¹⁾. The comprehensive analysis of metabolites in biological fluids using

Abbreviations: MEDE, METabolomics to characterise Dietary Exposure.

***Corresponding author:** J. Draper, fax +44 (0) 1970 621981, email jhd@aber.ac.uk

Table 1. Recent publications in which non-targeted metabolite fingerprinting or metabolite profiling has been used to discover potential dietary biomarkers in food intervention and cohort studies

Dietary exposure	Metabolomic technique	Potential biomarker-leads	Reference
<i>Acute food exposure studies (up to 24 h)</i>			
Chicken and orange juice	FIE-MS	Chicken: anserine; orange juice: proline betaine and 4-hydroxy-proline betaine	20
Apple juice	HPLC-ESI-MS/MS	Phloretin <i>O</i> -glucuronides and eight (methyl) quercetin <i>O</i> -glucuronides, 5-caffeoylquinic acid, 4- <i>p</i> -coumaroylquinic acid, caffeic acid, epicatechin, phloretin and quercetin	22
Cocoa powder with milk or water	HPLC-Q-ToF	Caffeine and theobromine (epicatechin, methylepicatechin, vanillic acid, vanilloylglycine, phenylvaleric acid, phenylvalerolactone derivatives, 3,5-diethyl-2-methylpyrazine and hydroxyacetophenone) and processing-derived products (diketopiperazines), as well as trigonelline, hydroxynicotinic acid, nicotinic acid and tyrosine	23
Cocoa powder and milk	HPLC-Q-ToF	Methylguanine, vanilloylglycine, dihydroxyphenyl valerolactone glucuronide, furoylglycine, 3- and 7-methylxanthine, theobromine and xanthurenic acid	24
Oily fish (salmon); a cruciferous vegetable (steamed broccoli); berry fruit (raspberries)	FIE-MS	Salmon: trimethylamine- <i>N</i> -oxide, methylhistidine, anserine. Raspberry: caffeoyl sulphate, methyl epicatechin sulphate, ascorbate, 3 hydroxylhipperic acid, naringenin glucuronide. Broccoli: ascorbate and breakdown products (xylonate/lyxonate, threitol/erythritol) naringenin and hesperitin glucuronide	25
Coffee	HPLC-PDA-MS	Caffeic acid sulphate, caffeoylquinic acid lactone sulphate, caffeoylquinic acid sulphate, dihydro(iso)ferulic acid glucuronide, dihydrocaffeic acid and glucuronide/sulphate, dihydroferulic acid and glucuronide/sulphate, feruloylglycine, feruloylquinic acid, iso/ferulic acid glucuronide/sulphate	32
<i>Short-term food interventions (up to 2 weeks)</i>			
Increased milk diet or an increased meat diet	¹ H NMR	Milk diet: increased hippurate; Meat diet: increased creatine, histidine and urea	17
High cruciferous vegetable (broccoli and Brussels sprouts)	¹ H NMR	<i>S</i> -methyl- <i>L</i> -cysteine sulphoxide (SMCSO) and three structurally related compounds	19
Mixed-fruit meal (apple, orange, grapes and grapefruit)	¹ H NMR	Citrus fruits: hippuric acid, proline betaine and tartaric acid	21
Dark chocolate	¹ H NMR and LC-MS	Increased levels of 4-hydroxyphenylacetate and several unassigned metabolites	28
'Vegetarian', 'low meat' and 'high meat' diets	¹ H NMR	High-meat diet: creatine, carnitine, acetylcarnitine and trimethylamine- <i>N</i> -oxide. Low-meat diet and vegetarian diet signatures: <i>p</i> -hydroxyphenylacetate	33
Low-phytochemical diet followed by a standard phytochemical diet	¹ H NMR	Hippurate increased in high phytochemical diet	35
<i>Long-term food interventions (> 1 month)</i>			
Wheat bread v. rye bread	HPLC-Q-ToF	3-(3,5-dihydroxyphenyl)-1-propanoic acid sulphate, enterolactone glucuronide, azelaic acid, 2-aminophenol sulphate and 2,4-dihydroxy-1,4-benzoxazin-3-one	18
High protein (HP) or low protein (LP) diet	¹ H NMR	Increased nitrogen, creatinine, TMAO, creatine after the HP diet	31
High mixed nuts consumption	HPLC-Q-ToF-MS	10-hydroxy-decene-4,6-dienoic acid sulphate, tridecadienoic/tridecynoic acid glucuronide, dodecanedioic acid, pyrogallol sulphate, <i>p</i> -coumaryl alcohol glucuronide, urolithin A glucuronide/sulphate/sulphoglucuronide, <i>N</i> -acetylserotonin sulphate and hydroxyindoleacetic acid	34



Table 1 (Continued)

Dietary exposure	Metabolomic technique	Potential biomarker-leads	Reference
<i>Habitual diet records used to classify individuals by intake of specific foods</i>			
Habitual diet examined to identify high, medium and low level consumers of citrus foods in a group of twenty-four participants	FIE-MS	Proline betaine and 4-hydroxy-proline betaine	26
Analysed diet records (FFQ) of free-living individuals (sixty-eight) and classified into high v. low consumers for a range of foods	FIE-MS	Markers for oily fish, orange juice, coffee and tomato described	27
Cohort diet diary data analysed for level of citrus consumption and eighty individuals selected for analysis (forty high citrus v. forty low citrus)	LC-MS	Proline betaine, flavanone glucuronides and limonene 8,9-diol glucuronide	30
Dietary patterns assessed using a 3 d diet diary in 160 individuals	NMR	High red meat: acetyl carnitine; High vegetable: phenylacetyl glutamine	29

HPLC-ESI-MS/MS, HPLC-electrospray ionisation-tandem MS; HPLC-PDA-MS, HPLC-photodiode array-MS; HPLC-Q-ToF, HPLC-quadrupole time-of-flight; FIE-MS, flow infusion electrospray-ionisation MS; LC-MS/MS, liquid chromatography tandem MS.

metabolomics technology provides an objective approach for the discovery of dietary exposure biomarkers^(11–35). Non-targeted metabolite fingerprinting, using either NMR^(17,19,21,28,29,31,33,35) or MS^(20,25–27), and metabolite profiling using liquid chromatography MS^(18,22–24,28,30,32,34) have been used successfully for biomarker-lead discovery using urine samples from various study designs (Table 1). In acute food intervention studies^(20,22–25,32), participants were exposed to specific foods in known amounts and either postprandial urines sampled before the next meal or overnight or 24 h urines were collected. In other studies, participants were established on a specific diet for several days/weeks^(17,19,21,28,33,35) or longer term (>1 month)^(18,31,34) before urine sampling. More recently there have been reports of the use of cohort studies in which participants consumed a freely-chosen diet^(26,27,29,30). In these studies, the analysis of diet diary or FFQ information allowed classification of individuals in terms of their frequency of consumption of specific diet constituents. In the present paper, we illustrate these approaches for biomarker discovery with particular reference to two studies carried out by the authors^(20,25–27).

The presence of substantial inter- and intra-individual variability in human metabolite profiles⁽³⁶⁾ provides a challenge for both biofluid sampling and subsequent data normalisation in metabolomics studies seeking information on habitual diet. To address this problem, standardised methods have been validated recently both for the management of participants and for urine sampling in large-scale food interventions involving free-living individuals⁽³⁷⁾ and also for acute postprandial studies in a controlled environment^(20,25). Key features of these study protocols include behavioural restrictions, e.g. no alcohol and the consumption of a standardised evening meal in the evening before a clinic visit to provide a fasting urine sample. It was anticipated that the latter would provide a ‘normalised’ background against which differences in urine chemistry resulting from either previous habitual dietary intake prior to clinic visit or acute food intake during the test day would be detectable^(12,14,20,25).

With effective protocols in place for volunteer management and urine sampling there was now an opportunity to determine whether changes in urine chemistry could reflect dietary exposure⁽¹¹⁾. In an acute feeding ‘proof of principle’ study, urine samples were analysed from individuals participating in the METabolomics to characterise Dietary Exposure (MEDE) research programme⁽²⁰⁾. As part of the MEDE project⁽¹¹⁾, twenty-four healthy participants consumed a ‘test’ breakfast, in which the cereal component of a standardised breakfast was replaced by one of four foods of high public health importance, followed by the collection of postprandial urine samples for metabolome analysis^(20,25). Once candidate food biomarkers had been identified^(25,26) there was then the opportunity to validate their potential usefulness to monitor habitual diet in the independent GrainMark study (<http://www.ncl.ac.uk/afrd/research/project/2287>)⁽²⁷⁾. This large-scale dietary intervention study, involving free-living individuals, aimed to discover potential biomarkers of dietary wholegrain exposure. After a washout period of 4 weeks the participants (sixty-eight in total) were asked to consume three servings of

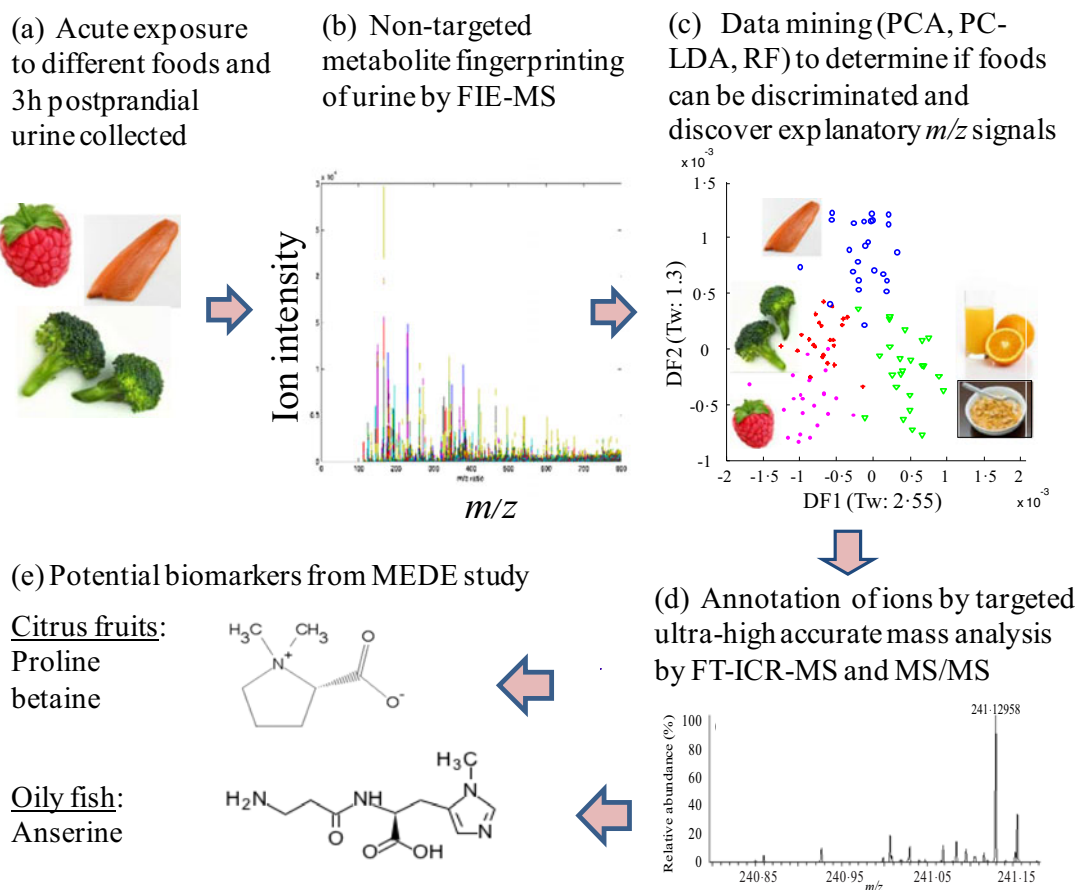


Fig. 1. (colour online) Typical analytical strategy for food biomarker discovery based on an acute food intervention. All elements of this example workflow are derived from data from experiments previously reported⁽²⁵⁾. (a) The standardised breakfast consisted of orange juice, a cup of tea with milk and sugar, a butter croissant and cornflakes with milk, while in the test breakfasts the cornflakes and milk were replaced with smoked salmon, steamed broccoli or raspberries. Postprandial urine was collected. (b) Typical flow infusion electrospray-ionisation MS (FIE-MS) fingerprints of postprandial urine. (c) Scores plot from a typical principal component-linear discriminant analysis (PC-LDA) of the metabolite fingerprinting data (m/z 100–800) derived from analysis postprandial urine. (d) Fourier transform-ion cyclotron resonance ultra-mass-spectrometry (FT-ICR-MS) plot of the nominal mass 'bin' m/z 241. (e) Potential biomarkers from the METabolomics to characterise Dietary Exposure (MEDE) study. PCA, principal component analysis; RF, random forest; MS/MS, flow infusion electrospray-ionisation tandem MS.

either wholegrain rye foods or wholegrain wheat foods per d for 4 weeks and subsequently doubled their intake of the same foods for a further 4 weeks. At baseline, and at the middle of each 4-week intervention period (washout, three servings and six servings of wholegrain rye/wheat foods per d), volunteers completed a validated FFQ (four in total) based on the European Prospective Investigation into Cancer and Nutrition FFQ⁽⁷⁾, which recorded consumption of foods for a 7 d period, within 7–14 d of sampling⁽³⁷⁾. Using data from both the MEDE and GrainMark studies it has been shown that analysis of overnight void urine samples can provide a rich source of potential biomarkers of habitual diet as reported in FFQ^(26,27).

Postprandial urine composition reflects recent dietary exposure

Table 1 lists several recent studies in which non-targeted metabolite fingerprinting or metabolite profiling has been

used to discover dietary biomarkers in both acute and short-term food intervention studies^(17,19–25,28,32,33,35). The basic design principles are illustrated in Fig. 1 with reference to the MEDE study in which fasted participants consumed specific foods (Fig. 1(a)) as part of a standardised breakfast. Metabolome fingerprints representing postprandial urines⁽²⁵⁾ were then generated using non-targeted, nominal mass flow injection electrospray-ionisation MS^(15,20,25). Figure 1(b) illustrates a typical flow injection electrospray-ionisation MS urine fingerprint of a mass range from m/z 100 to 800, which shows that urine fingerprints are both complex and information-rich. The question of whether a postprandial urine sample contains chemicals distinctive of exposure to specific foods can be evaluated by subjecting the urine fingerprint data to powerful supervised multivariate analysis including principal component-linear discriminant analysis. Figure 1(c) shows a scores plot from a typical principal component-linear discriminant analysis⁽¹⁶⁾ of flow injection

electrospray-ionisation MS fingerprints representing urine samples from volunteers exposed to a standard breakfast or to a breakfast in which the cereal component of the standard breakfast was replaced by smoked salmon, broccoli or raspberries⁽²⁵⁾. With an eigenvalue (Tw) of 2.55 in the dimension of maximal discrimination (Fig. 1(c)), it is evident that the 'test' foods (particularly smoked salmon) are adequately discriminated from the standard breakfast. A range of feature selection methods⁽¹⁶⁾ can then be employed to determine the flow injection electrospray-ionisation MS signals responsible for the discrimination of each food and, under most circumstances, it has been found that agglomerative decision trees such as random forest perform consistently well^(20,25,26,38). The detailed analysis of highlighted nominal masses can be undertaken by targeting them for further investigation on MS instruments capable of ultra-high mass resolution^(25–26). The Fourier transform-ion cyclotron resonance ultra-MS plot shown in Fig. 1(d) represents the analysis of the nominal mass 'bin' m/z 241, a signal linked with the consumption of oily fish, indicating that a signal with a mass of 241.12958 is a likely biomarker candidate for oily fish exposure⁽²⁵⁾. The accurate mass information can be used to generate an elemental formula and to predict candidate metabolites that could yield the measured ion using annotation tools such as MZedDB⁽³⁹⁾, which take into account anticipated ionisation behaviour. Further experiments in which the selected ion is fragmented and the resulting spectrum compared with that of a chemical standard are required to assign a putative structure. For example, in the MEDE study these subsequent targeted analyses suggested that anserine (nominal mass m/z 241) is a candidate biomarker for oily fish consumption and proline betaine for citrus foods (Fig. 1(e))^(25,26).

Exposure to citrus foods has provided a paradigm for discovery of dietary biomarkers by analysis of dietary data

Citrus fruits and citrus fruit juices are a distinctive and frequently consumed (i.e. often once per d) component of the UK diet and thus represent a paradigm for validation of biomarkers of habitual dietary exposure. Studies using either NMR⁽²¹⁾ or ESI-MS fingerprinting⁽²⁰⁾ of post-prandial urines demonstrated initially that proline betaine was a potential biomarker of acute exposure to citrus foods. There was thus an opportunity to determine whether the urinary concentration of this metabolite reflected habitual exposure to the same foods. Recent publications described a multivariate modelling strategy for habitual dietary biomarker discovery dependent on the comparison of urine metabolome fingerprints from groups of individuals reporting differential exposure to specific dietary components^(26,27,29,30). An initial analysis⁽²⁶⁾ of habitual citrus exposure data in the MEDE study (n 24) indicated that there were sufficient individuals reporting citrus consumption to allow the assignment of volunteers into three broad exposure levels (High (two–three citrus portions per d), Medium (about one citrus portion per d) and Low (less than two citrus portions per week)). With the larger

number of participants in the GrainMark study it was possible to assign individuals to a more quantitative scale of habitual dietary exposure (i.e. never, 1 per week, 2–4 per week, 5–6 per week, 1 per d, 2–3 per d)⁽²⁷⁾. Overnight void urine samples available from both studies were subjected to metabolite fingerprinting and the m/z signals responsible for discriminating higher *v.* lower habitual consumption levels were identified using random forest feature selection⁽¹⁶⁾. In both studies the majority of the top twenty highest ranked signals responsible for the classification of *High v. Low* habitual citrus consumption were identified as ionisation products of proline betaine^(26,27). Although not strictly quantitative, the relative intensity of one of the major ions (m/z 144; [M+H]⁺) reflected the level of habitual citrus exposure reported by individual participants in both MEDE and GrainMark studies. A similar result has been reported for NMR signals associated with recent exposure to citrus foods that were found to be derived from the presence of proline betaine in 24 h urine samples associated with recent exposure to citrus foods⁽²¹⁾.

Habitual consumption frequency for individual foods impacts on the ability to detect differential dietary exposure using overnight urine samples

Following the success with citrus foods there have been several recent reports in which dietary data have been analysed from studies where, in most cases, participants have eaten freely-chosen diets (Table 1)^(26,27,29,30). These meta-data allowed the assignment of individuals into different food consumption frequency classes and could be used to study habitual exposure for a range of other foods. In the GrainMark study, as expected, habitual consumption frequency differed greatly between individual foods, and these food-specific consumption patterns were generally consistent with the MEDE study⁽²⁷⁾. Overall, the patterns of intake of individual foods could be summarised in four general exposure categories (Fig. 2) ranging from foods consumed very infrequently (e.g. liver or kidneys) to those consumed, on average, more than once per d (e.g. coffee). Although habitual citrus exposure could be modelled adequately using data from the MEDE study⁽²⁶⁾, with only twenty-four participants the present strategy was not suitable for biomarker discovery for foods which are consumed very infrequently by most people. Assignment to habitual consumption frequency ranges for the remaining foods (grouped into High skewed, Normal distribution or Low skewed consumption patterns) identified sufficient numbers of individuals within the GrainMark study to develop higher *v.* lower frequency exposure groups for multivariate classification⁽²⁷⁾. The likelihood of discovering potential biomarkers for each food was assessed by determining the 'goodness' of class discrimination using random forest margin values and area under the receiver operating characteristic curve values as robust classification statistics⁽¹⁶⁾. Figure 2 illustrates a clear trend that classification efficiency (as assessed by random forest margins and area under the receiver operating characteristic curve values) is generally higher in foods that are consumed

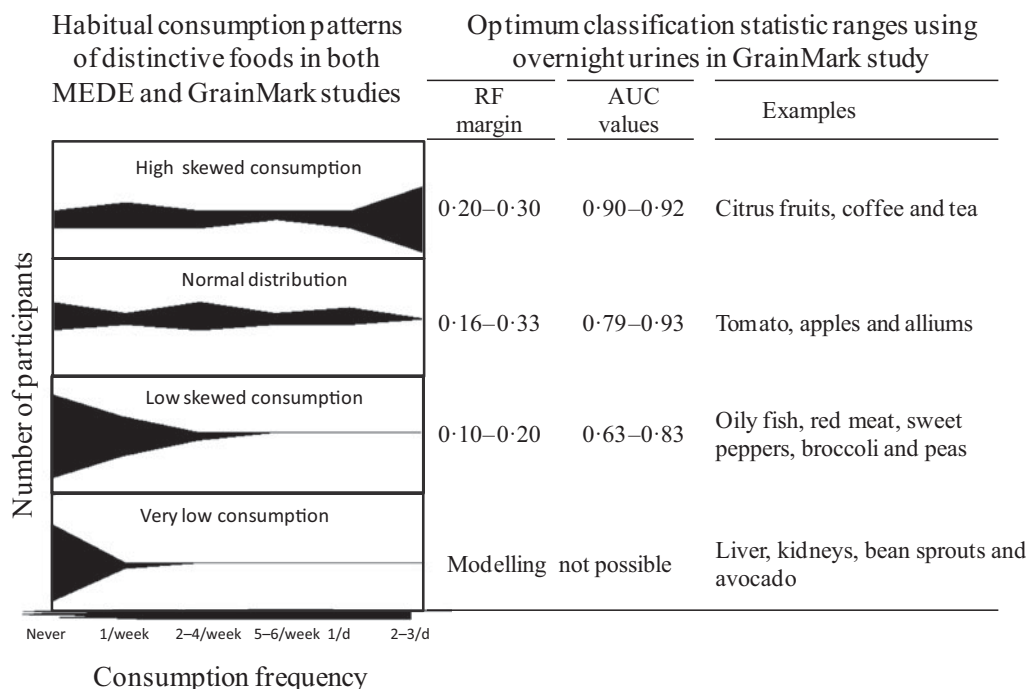


Fig. 2. Habitual food consumption patterns and overnight urine fingerprint modelling outcomes. Consumption patterns in both the MEtabolomics to characterise Dietary Exposure (MEDE) and GrainMark studies^(26–27). Statistic ranges from the GrainMark study in positive ionisation mode. RF margin, random forest margin values; AUC, area under the receiver operating characteristic curve. Figures are derived from the data from experiments reported previously^(26,27).

more frequently. We suggest that such foods will form likely candidates for monitoring using a biomarker strategy if it can be proven that the presence of potential biomarker signals in urine can be linked to the ingestion and chemical composition of specific dietary components. An example ‘data-driven’ strategy for the discovery of potential dietary biomarkers based on the analysis of GrainMark FFQ structure and data is summarised in Fig. 3.

Structural analysis of potential biomarkers representing three selected foods shows that their presence in overnight void urine reflects original food chemical composition

Metabolome models (based on overnight void urine samples) of three distinctive foods representing examples of Low skewed, Normal distribution and High skewed habitual dietary exposure groups have been investigated in more detail to determine whether the selected metabolites could be correlated with known food chemistry^(10,26,32,33,40–51). Groups of participants were identified for each dietary component to represent High, Medium and Low exposure classes within the consumption ranges described for each food in the GrainMark study. Thus for oily fish (Low skewed) the Low consumption category represents less than one portion per week, Medium is approximately one per week and High is greater than or equal to two to four portions week. For tomato (Normal distribution), the Low consumers ate less than or equal to one portion per week, Medium consumers ate greater than two to four

portions per week, while High consumers ate up to five to six portions per week. High consumers of coffee (High skewed) drank more than one cup per d while Medium consumers had greater than or equal to two to four cups per week and Low consumers more than or equal to one per week. Despite the wide range of consumption frequencies each of these distinctive foods generated adequate classification models (Fig. 4(a)).

In the GrainMark study, dihydrocaffeic acid-3-*O*-glucuronide was highly ranked as a potential biomarker of habitual coffee exposure (Fig. 4(b)). Previous studies on acute exposure to coffee reported the presence in urine and/or plasma of at least ten phenolic compounds which represented metabolic endpoints of the biotransformation of chlorogenic acid and caffeic acid which is present at high levels in this beverage^(10,32,43,49). Although chlorogenic acid and caffeic acid are found in many fruits and vegetables and it is unlikely that dihydrocaffeic acids will prove to be unique biomarkers of coffee consumption, it is interesting that other phenolic metabolites identified in acute exposure studies were not highlighted by this analysis. Investigation of top ranked signals for oily fish exposure showed that methyl-histidine (probably derived from anserine, which was also highly ranked, by the action of carnosinase)^(40–42), was an excellent biomarker candidate (Fig. 4(c)). Thus, the identity of both of these potential biomarkers could be linked directly to the previous chemical analysis of each food. Although hippuric acids (presumably derived from colonic fermentation of the hydroxycinnamic acid content of tomato fruits⁽⁴⁸⁾) were indicative of tomato exposure, many highly ranked signals

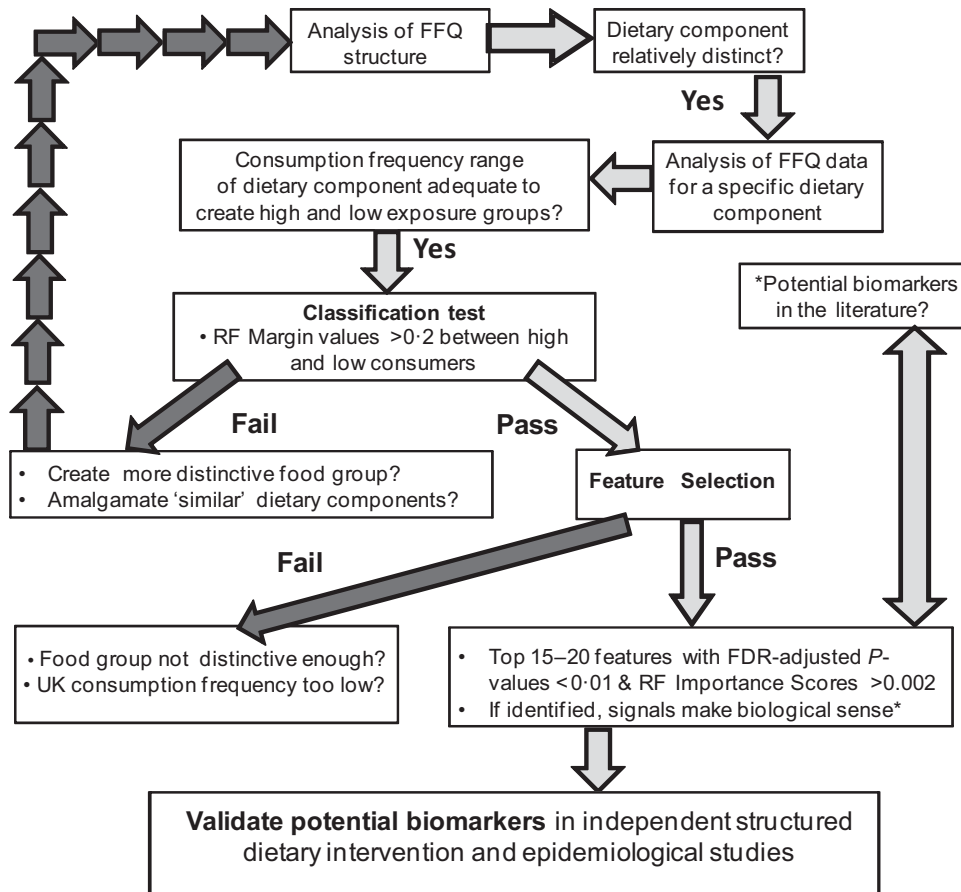


Fig. 3. Example strategy for 'data-driven' biomarker discovery based on the GrainMark study using overnight urine samples. RF, random forest; FDR, false discovery rate; dark grey arrows denote fail; light grey arrows denote pass. Figure adapted from that presented previously⁽²⁷⁾.

proved to be dihydroxyphenylvalerolactone conjugates that commonly result from the colonic fermentation of flavonoids, such as flavanols⁽⁵²⁾ (Fig. 4(d)). As tomato contains insignificant amounts of such polyphenols it is most likely that these compounds are associated with foods strongly co-consumed in meals containing tomatoes. This observation highlights the requirement for careful validation of any potential food biomarkers in the context of the whole diet and emphasises the utility of the test meal approach (as used in the MEDE study) for the identification of putative biomarkers which are causally linked with food exposure. In addition, such data-driven strategies for putative biomarker discovery are constrained by the robustness of the original dietary intake data which may contain inherent biases. Further, both conceptual and practical challenges are likely in cases where there are strong correlations between intakes of foods with related chemistries.

Conclusion

The advent of non-targeted metabolomics technology for global chemical fingerprinting/profiling of human biofluids has offered an opportunity to accelerate research on food biomarker discovery. Recent data support the concept that metabolomics analysis of 24 h or overnight void urine

samples, in particular, will provide a productive strategy for the identification of candidate dietary exposure biomarkers. The demonstration that biomarker discovery using high throughput metabolomics is feasible using urine samples derived from populations of free-living individuals^(26,27,29,30) supports the ultimate objective of using such biomarkers in epidemiological studies. Evidence is now accumulating that a range of relatively frequently consumed (i.e. one to two portions per week) and more distinctive foods can be considered good candidates for biomarker discovery using biofluid samples from cohort studies^(26,27,29,30). Many other foods forming major components of composite meals, or less frequently consumed food items, may be targets for biomarker discovery but would probably require well designed, controlled food intervention studies to identify candidate metabolites. In future studies we propose testing this hypothesis by the analysis of overnight urine samples collected at home by individuals consuming carefully constructed weekly menus designed both to provide adequate exposure to specific dietary components and to offer foods in specific combinations in order to expose any biomarker redundancy that could confound dietary exposure measurement.

Ensuring the generation of accurate dietary information in epidemiological studies requires considerable effort from both researchers and study participants. Despite the

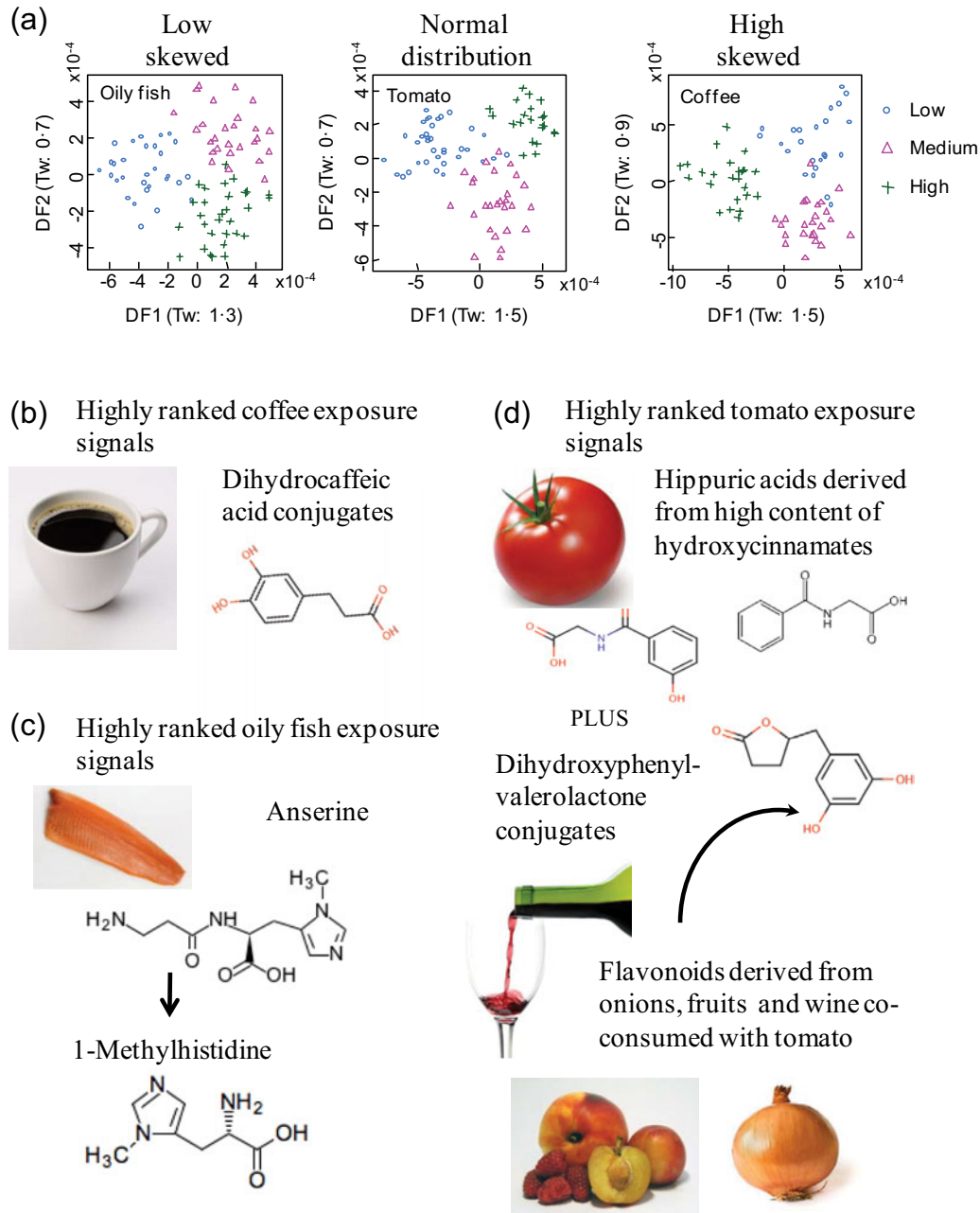


Fig. 4. (colour online) Habitual consumption of distinctive foods eaten in a range of consumption frequencies is detectable in overnight urine samples by flow infusion electrospray-ionisation MS (FIE-MS) fingerprinting. The data illustrated in this figure are derived from experiments previously reported⁽²⁷⁾. (a) Principal component-linear discriminant analysis (PC-LDA) of the metabolite fingerprinting data (m/z 100–550; oily fish, positive ionisation mode; tomato and coffee, negative mode) derived from the analysis of overnight ‘PRE’ urine from participants classified as low/medium/high exposure ranges (ten to twelve per exposure group) based on the average FFQ score for exposure to these different foods. Oily fish: low, less than one per week; medium, one per week; high, greater than or equal to two to four per week; tomato: low, less than or equal to one per week; medium, greater than two to four per week to less than five to six per week; high, greater than or equal to five to six per week; coffee: low, less than or equal to one per week; medium, greater than or equal to two to four per week to less than five to six per week; high, greater than one per week; eigenvalues (Tw values) are given in parentheses. Markers confirmed using Fourier transform-ion cyclotron resonance ultra-mass-spectrometry (FT-ICR-MS), flow infusion electrospray-ionisation tandem MS and MZedDB. (b) Signals highly ranked in habitual coffee exposure models. (c) Signals highly ranked in habitual oily fish exposure models. (d) Signals highly ranked in habitual tomato exposure models.

advent of increasingly sophisticated digital tools⁽¹⁾, diet recording remains an inherent source of major uncertainty, even in studies where individuals recording food exposures are well trained and carefully monitored. Following validation in epidemiological and/or controlled dietary intervention studies, it is expected that putative food intake biomarkers can be translated into practical measurements that can complement, or in some cases replace, more traditional methods of assessing dietary exposure. We conclude that in the not too distant future, urine biomarker technology may allow objective monitoring of the levels of intake of several key foods and strengthen the evidence for causal links between dietary exposure and health outcomes.

Acknowledgements

None of the authors has a conflict of interest with respect to this manuscript. This research programme was supported by the UK Food Standards Agency Projects N05073 and N05075 and MRC Programme Grant MR/J010308/1. The authors' contributions to the work were as follows: M. B. developed urine extraction procedures, designed metabolite fingerprinting experiments, supervised MS support staff, pre-processed data for analysis and edited the manuscript; A. J. L. was involved in data analysis, produced figures, researched the literature and wrote the manuscript; G. F. and S. H. undertook volunteer recruitment, coordinated volunteer Clinical Research Facility visits and supervised Clinical Research Facility support staff and edited the manuscript; J. C. M., C. S. and K. B. coordinated the project, supervised research in Newcastle University, designed volunteer handling protocols and edited the manuscript; and J. D. coordinated the project, supervised research in Aberystwyth, designed figures and wrote the manuscript.

References

1. Penn L, Boeing H, Boushey CJ *et al.* (2010) Assessment of dietary intake: NuGO symposium report. *Genes Nutr* **5**, 205–213.
2. Bingham SA, Gill C, Welch A *et al.* (1994) Comparison of dietary assessment methods in nutritional epidemiology – weighed records *v* 24-h recalls, food-frequency questionnaires and estimated-diet records. *Br J Nutr* **72**, 619–643.
3. Kristal AR, Peters U & Potter JD (2005) Is it time to abandon the food-frequency questionnaire? *Cancer Epidemiol Biomarkers Prev* **14**, 2826–2828.
4. Holmes MD, Powell IJ, Campos H *et al.* (2007) Validation of a food-frequency questionnaire measurement of selected nutrients using biological markers in African–American men. *Eur J Clin Nutr* **61**, 1328–1336.
5. Cade J, Thompson R, Burley V *et al.* (2002) Development, validation and utilisation of food-frequency questionnaires – a review. *Public Health Nutr* **5**, 567–587.
6. Jia X, Craig LCA, Aucott LS *et al.* (2008) Repeatability and validity of a food-frequency questionnaire in free-living older people in relation to cognitive function. *J Nutr Health Aging* **12**, 735–741.
7. Bingham SA, Gill C, Welch A *et al.* (1997) Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. *Int J Epidemiol* **26**, S137–S151.
8. Jenab M, Slimani N, Bictash M *et al.* (2009) Biomarkers in nutritional epidemiology: applications, needs and new horizons. *Hum Genet* **125**, 507–525.
9. Signorello LB, Buchowski MS, Cai Q *et al.* (2010) Biochemical validation of food-frequency questionnaire-estimated carotenoid, alpha-tocopherol, and folate intakes among African–Americans and non-Hispanic whites in the Southern Community Cohort Study. *Am J Epidemiol* **171**, 488–497.
10. Mennen LI, Sapinho D, Ito H *et al.* (2006) Urinary flavonoids and phenolic acids as biomarkers of intake for polyphenol-rich foods. *Br J Nutr* **96**, 191–198.
11. Favé G, Beckmann ME, Draper J *et al.* (2009) Measurement of dietary exposure: a challenging problem which may be overcome thanks to metabolomics? *Genes Nutr* **4**, 135–141.
12. Scalbert A, Brennan L, Fiehn O *et al.* (2009) Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics* **5**, 435–458.
13. Holmes E, Loo RL, Stalmer J *et al.* (2008) Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* **453**, 396–400.
14. Walsh MC, Brennan L, Malthouse JPG *et al.* (2006) Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. *Am J Clin Nutr* **84**, 531–539.
15. Beckmann M, Parker D, Enot DP *et al.* (2008) High-throughput, nontargeted metabolite fingerprinting using nominal mass flow injection electrospray mass spectrometry. *Nat Protoc* **3**, 486–504.
16. Enot DP, Lin W, Beckmann M *et al.* (2008) Preprocessing, classification modeling and feature selection using flow injection electrospray mass spectrometry metabolite fingerprint data. *Nat Protoc* **3**, 446–470.
17. Bertram HC, Hoppe C, Petersen BO *et al.* (2007) An NMR-based metabolomic investigation on effects of milk and meat protein diets given to 8-year-old boys. *Br J Nutr* **97**, 758–763.
18. Bondia-Pons I, Barri T, Hanhineva K *et al.* (2013) UPLC-QTOF/MS metabolic profiling unveils urinary changes in humans after a whole grain rye versus refined wheat bread intervention. *Mol Nutr Food Res* **57**, 412–422.
19. Edmands WMB, Beckonert OP, Stella C *et al.* (2011) Identification of human urinary biomarkers of cruciferous vegetable consumption by metabolomic profiling. *J Proteome Res* **10**, 4513–4521.
20. Favé G, Beckmann M, Lloyd AJ *et al.* (2011) Development and validation of a standardized protocol to monitor human dietary exposure by metabolite fingerprinting of urine samples. *Metabolomics* **7**, 469–484.
21. Heinzmann SS, Merrifield CA, Rezzi S *et al.* (2012) Stability and robustness of human metabolic phenotypes in response to sequential food challenges. *J Proteome Res* **11**, 643–655.
22. Kahle K, Kempf M, Schreier P *et al.* (2011) Intestinal transit and systemic metabolism of apple polyphenols. *Eur J Nutr* **50**, 507–522.
23. Llorach R, Urpi-Sarda M, Jauregui O *et al.* (2009) An LC-MS-based metabolomics approach for exploring urinary metabolome modifications after cocoa consumption. *J Proteome Res* **8**, 5060–5068.
24. Llorach-Asuncion, R, Jauregui O, Urpi-Sarda M *et al.* (2010) Methodological aspects for metabolome visualization and characterization: a metabolomic evaluation of the 24h



- evolution of human urine after cocoa powder consumption. *J Pharmaceut Biomed* **51**, 373–381.
25. 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**, 981–991.
 26. Lloyd AJ, Beckmann M, Favé G *et al.* (2011) Proline betaine and its biotransformation products in fasting urine samples are potential biomarkers of habitual citrus fruit consumption. *Br J Nutr* **106**, 812–824.
 27. Lloyd AJ, Beckmann M, Haldar S *et al.* (2013) Data-driven strategy for the discovery of potential urinary biomarkers of habitual dietary exposure. *Am J Clin Nutr* **97**, 377–389.
 28. Martin F-PJ, Rezzi S, Peré-Trepat E *et al.* (2009) Metabolic effects of dark chocolate consumption on energy, gut microbiota, and stress-related metabolism in free-living subjects. *J Proteome Res* **8**, 5568–5579.
 29. O'Sullivan A, Gibney MJ & Brennan L (2011) Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. *Am J Clin Nutr* **93**, 314–321.
 30. Pujos-Guillot E, Hubert J, Martin JF *et al.* (2013) Mass spectrometry-based metabolomics for the discovery of biomarkers of fruit and vegetable intake: citrus fruit as a case study. *J Proteome Res* **12**, 1645–1659.
 31. 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**, 112–131.
 32. Stalmach A, Mullen W, Barron D *et al.* (2009) Metabolite profiling of hydroxycinnamate derivatives in plasma and urine after the ingestion of coffee by humans: identification of biomarkers of coffee consumption. *Drug Metabol Dispos* **37**, 1749–1758.
 33. Stella C, Beckwith-Hall B, Cloarec O *et al.* (2006) Susceptibility of human metabolic phenotypes to dietary modulation. *J Proteome Res* **5**, 2780–2788.
 34. Tulipani S, Llorach R, Jáuregui O *et al.* (2011) Metabolomics unveils urinary changes in subjects with metabolic syndrome following 12-week nut consumption. *J Proteome Res* **10**, 5047–5058.
 35. Walsh MC, Brennan L, Pujos-Guillot E *et al.* (2008) Influence of acute phytochemical intake on human urinary metabolomic profiles. *Am J Clin Nutr* **86**, 1687–1693.
 36. Assfalg M, Bertini I, Colangiuli D *et al.* (2008) Evidence of different metabolic phenotypes in humans. *Proc Natl Acad Sci USA* **105**, 1420–1424.
 37. Brownlee IA, Moore C, Chatfield M *et al.* (2011) Markers of cardiovascular risk are not changed by increased whole grain-intake. The WHOLEheart study: a randomised, controlled dietary intervention. *Br J Nutr* **104**, 125–134.
 38. Beckmann M, Enot DP, Overy DP *et al.* (2010) Metabolite fingerprinting of urine suggests breed-specific dietary metabolism differences in domestic dogs. *Br J Nutr* **103**, 1127–1138.
 39. Draper J, Enot DP, Parker D *et al.* (2009) Metabolite signal identification in accurate mass metabolomics data with MZedDB, an interactive *m/z* annotation tool utilising predicted ionisation behaviour 'rules'. *BMC Bioinf* **10**, 24.
 40. Abe H (1983) Distribution of free L-histidine and related dipeptides in the muscle of fresh-water fishes. *Comp Biochem Physiol Part B Biochem Mol Biol* **76**, 35–39.
 41. Abe H, Okuma E, Sekine H *et al.* (1993) Human urinary excretion of L-histidine-related compounds after ingestion of several meats and fish muscle. *Int J Biochem* **25**, 1245–1249.
 42. Dragsted LO (2010) Biomarkers of meat intake and the application of nutrigenomics. *Meat Sci* **84**, 301–307.
 43. Ito H, Gonthier MP, Manach C *et al.* (2005) Polyphenol levels in human urine after intake of six different polyphenol-rich beverages. *Br J Nutr* **94**, 500–509.
 44. Jacob K, Periago MJ, Boehm V *et al.* (2008) Influence of lycopene and vitamin C from tomato juice on biomarkers of oxidative stress and inflammation. *Br J Nutr* **99**, 137–146.
 45. Jaganath IB, Mullen W, Edwards CA *et al.* (2006) The relative contribution of the small and large intestine to the absorption and metabolism of rutin in man. *Free Radical Res* **40**, 1035–1046.
 46. Lee MB, Storer MK, Blunt JW *et al.* (2006) Validation of H-1 NMR spectroscopy as an analytical tool for methylamine metabolites in urine. *Clin Chim Acta* **365**, 264–269.
 47. Mahaffey KR (2004) Fish and shellfish as dietary sources of methylmercury and the omega-3 fatty acids, eicosahexaenoic acid and docosahexaenoic acid: risks and benefits. *Environ Res* **95**, 414–428.
 48. Sánchez-Rodríguez E, Ruiz JM, Ferreres F *et al.* (2012) Phenolic profiles of cherry tomatoes as influenced by hydric stress and rootstock technique. *Food Chem* **134**, 775–782.
 49. Rechner AR, Spencer JPE, Kuhnle G *et al.* (2001) Novel biomarkers of the metabolism of caffeic acid derivatives *in vivo*. *Free Radical Biol Med* **30**, 1213–1222.
 50. Vallverdu-Queralt A, Jauregui O, Medina-Remon A *et al.* (2010) Improved characterization of tomato polyphenols using liquid chromatography/electrospray ionization linear ion trap quadrupole orbitrap mass spectrometry and liquid chromatography/electrospray ionization tandem mass spectrometry. *Rapid Commun Mass Spectrom* **24**, 2986–2992.
 51. Visciano P, Perugini M, Manera M *et al.* (2009) Selected polycyclic aromatic hydrocarbons in smoked tuna, swordfish and Atlantic salmon fillets. *Int J Food Sci Technol* **44**, 2028–2032.
 52. Urpi-Sarda M, Monagas M, Khan N *et al.* (2009) Targeted metabolic profiling of phenolics in urine and plasma after regular consumption of cocoa by liquid chromatography-tandem mass spectrometry. *J Chromatogr A* **1216**, 7258–7267.