Use of mass spectrometry fingerprinting to identify urinary metabolites after consumption of specific foods\textsuperscript{1–4}

Amanda J Lloyd, Gaëlle Fave, Manfred Beckmann, Wanchang Lin, Kathleen Tailliart, Long Xie, John C Mathers, and John Draper

ABSTRACT

Background: The lack of robust biological markers of dietary exposure hinders the quantitative understanding of causal relations between diet and health.

Objective: We aimed to develop an efficient procedure to discover metabolites in urine that may have future potential as biomarkers of acute exposure to foods of high public health importance.

Design: Twenty-four participants were provided with a test breakfast in which the cereal component of a standardized breakfast was replaced by 1 of 4 foods of high public health importance: 1.5-, 3-, and 4.5-h postprandial urine samples were collected. Flow infusion electrospray–ionization mass spectrometry followed by supervised multivariate data analysis was used to discover signals resulting from consumption of each test food.

Results: Fasted-state urine samples provided a universal comparator for food biomarker lead discovery in postprandial urine. The filtering of data features associated with consumption of the common components of the standardized breakfast improved discrimination models and readily identified metabolites that showed consumption of specific test foods. A combination of trimethylamine-N-oxide and 1-methylhistidine was associated with salmon consumption. Novel ascorbate derivatives were discovered in urine after consumption of either broccoli or raspberries. Sulphonated caffeic acid and sulphonated methyl-epicatechin concentrations increased dramatically after consumption of raspberries.

Conclusions: This biomarker lead discovery strategy can identify urinary metabolites associated with acute exposure to individual foods. Future studies are required to validate the specificity and utility of potential biomarkers in an epidemiologic context.

INTRODUCTION

To strengthen the evidence base for dietary recommendations concerning the consumption of specific foods to reduce the risk of major chronic diseases (1–6) and to monitor the effects of dietary interventions, it is essential to have robust methods for measuring dietary exposure (7). Unfortunately, many existing dietary assessment methods are difficult to validate, are subject to participant bias, and depend on food-composition tables to estimate intakes of energy, nutrients, and other food constituents (7–9). As a consequence, there is a need for novel approaches for dietary exposure measurement, which offer greater objectivity and reliability (7). Recent studies described the use of targeted analysis, in blood and urine samples, of specific nutrients and metabolites as biomarkers of dietary intake (8, 10–13). Nontargeted metabolomics technology, which allows comprehensive monitoring of metabolites in biological fluids, provides an alternative nonsubjective approach to assessing food intake without any prior knowledge of specific food chemistry (14, 15). Metabolite fingerprinting methods such as NMR\textsuperscript{5} spectrometry (16, 17) and FIE-MS (14, 15, 18, 19) have great potential for use in large-scale studies.

As prerequisites for the development of metabolomics-based approaches for assessing dietary exposure, we validated a method for the management of study participants (14, 20) and for the collection, processing, and analysis of “behavioral phase” urine samples (20, 21). Many metabolites currently reported in the literature as potential biomarkers of habitual dietary exposure are absorbed in the upper gastrointestinal tract within 1.0–1.5 h of consumption (7, 8, 13, 14, 20–26) and first appear in urine within a further 1.5–2.5 h. We propose that acute phase postprandial urine may provide a valuable source of dietary exposure biomarker leads. After a standardized meal low in polyphenols in the evening before each experimental day, a standardized breakfast was consumed by fasting participants to provide a specific acute dietary challenge (20). With the use of FIE-MS (18) in conjunction with supervised multivariate data analysis (27) and MZedDB [an m/z signal annotation tool; (28)], we showed that postprandial urine samples collected 2–4 h after...
consumption of the standardized breakfast have a stable, reproducible, and information-rich metabolite composition (20).

In the current study, we describe an efficient procedure for potential biomarker lead discovery in postprandial urine samples by replacing the cereal component of the standardized breakfast (cornflakes with milk) with a panel of test foods of high public health importance. Using a data-filtering approach to remove signals resulting from consumption of the common components of the standardized breakfast, we showed that highly ranked signals discriminating postprandial from fasting urine samples represent metabolites derived from compounds known to be abundant in each test food.

SUBJECTS AND METHODS

Ethical approval and subject recruitment

This project was approved by the Newcastle and North Tyneside 2 Research Ethics Committee (reference number 07/H0907/136) and registered with the Newcastle on Tyne Hospitals NHS Foundation Trust, United Kingdom (registration number 4392). The project constitutes part of the MEDE research program (14), which aimed to develop a standardized protocol for nutritional metabolomics investigations (20) and involved 3 different studies known as MEDE Study 1, MEDE Study 2, and MEDE Study 3. Subject recruitment started in November 2007. In the current report, we investigated MEDE Study 2 participants (n = 12) and MEDE Study 3 participants (n = 24) (14). The study participants were assessed for suitability via a screening questionnaire (see reference 20 for further details) and were recruited at an induction visit to the CRF (Royal Victoria Infirmary). Biodata for MEDE Study 2 and MEDE Study 3 participants are presented elsewhere (see Supplemental data 1 and 2, respectively, under “Supplemental data” in the online issue).

Study design

In line with our previous research (14, 20, 21), we imposed behavioral constraints on study participants (minimal physical activity and no alcohol consumption on the day before the test day) and provided a standardized evening meal (ready-prepared meal consisting of a chicken breast with carrots, peas, and roasted potatoes in an onion gravy, one 75-g chocolate éclair, and a 500-mL bottle of still mineral water to be consumed between 1800 and 2000 h for full details of all food items; 20). Study participants were provided with a second bottle of still mineral water to be consumed ad libitum until arrival at the CRF on the morning of the study day. On each study day, participants came to the CRF after a 12-h minimum fast, and fasting urine samples were collected. MEDE Study 2 participants came to the CRF on 2 occasions and consumed an SB (orange juice, tea with skimmed milk and sugar, butter croissant, and cornflakes with milk; 20); all visits took place over a 6-wk period in the spring/early summer of 2008. MEDE Study 3 participants visited the CRF 6 times over an 8-mo period from September 2008 to April 2009, and each were provided with either the SB or with a breakfast in which the corn flakes and milk of the SB were replaced with a test food of high public health importance (Figure 1). The 4 test foods consisted of oily fish (60 g smoked salmon trimmings), a cruciferous vegetable (200 g steamed broccoli florets), a berry fruit (200 g raspberries), or a whole-grain wheat cereal (37.5 g; 2 biscuits) with 125 mL ultra-high-temperature–treated semiskimmed milk. Urine samples were collected 1.5, 3, and 4.5 h after consumption of the breakfast. Each of the 24 MEDE Study 3 participants consumed the SB twice and each of the 4 tests once. The order of investigation of each of the 6 meals for random groups of 6 volunteers was decided according to a 6 × 6 Latin-Square design in which each test meal occurred once and each SB twice in each row (order of presentation) with the order of presentation unique to each column (study participant). This was repeated with unique Latin Squares for additional groups of 6 study participants to result in a study design with four 6 × 6 Latin Squares.

The degree of replication was designed to allow sufficient statistical power to construct robust classification as reported previously (27). Urine samples were frozen immediately at –20° C and then at –80° C within 24 h (see reference 20 for full details of all laboratory protocols and consumables).

FIGURE 1. Schematic representation of the design of MEDE Study 3. On the pretest day, participants were asked to empty their bladder and discard the urine before consuming their standardized evening meal between 1800 and 2000 h. Study participants (n = 24) collected all urine produced after consumption of the evening meal, up to and including the morning void, before attending the clinical research facility (“PRE” sample). On the morning of the test day, fasting urine samples [after a 12-h (minimum) fast] were collected, and then the breakfast was served. The standardized breakfast consisted of orange juice, a cup of tea with milk and sugar, a butter croissant, and cornflakes with milk. For the test breakfasts, the cornflakes and milk in the standardized breakfast were replaced with smoked salmon, steamed broccoli, raspberries, or whole-grain wheat cereal with milk. Postprandial urine samples were collected 1.5, 3, and 4.5 h after the participants had consumed each breakfast. MEDE, MEtabolomics to characterize Dietary Exposure.
FIE-MS

FIE-MS was carried out as described previously (18–21). Aliquots of thawed urine (50 μL) were diluted in 450 μL pre-chilled methanol:water (3.5:1), vortex mixed, shaken for 15 min at 4°C, and then centrifuged for 5 min at 14,000 × g. Data were acquired in alternating positive and negative ionization modes and over 4 scan ranges (15–110, 100–220, 210–510, and 500–1200 m/z), with an acquisition time of 5 min, on an LTQ linear ion trap (Thermo Electron Corporation). The resulting mass spectrum was the mean of 20 scans about the apex of the infusion profile. Raw data dimensionality was reduced by electronically extracting signals with ± 0.1-Da mass accuracy and (unless shown otherwise) are presented at 1-amu accuracy in the figures and tables. Data were \( \log_{10} \) transformed and normalized to total ion current before analysis (18).

GC-tof-MS

GC-tof-MS analysis was carried out as described previously (19); 50 μL of 0.1 mmol internal standard/L [\( \ominus \)-three-furfurylserine dissolved in methanol:water (70:30)] was added to 300 μL of the urine supernatant fluid. Samples were then dried in a speed vacuum. Two-step derivatization of dried samples was carried out by protecting the carbonyl moieties by methoximation with the use of 60 μL of a 20-mg/mL solution of methoxyamine hydrochloride (Fluka) in pyridine (Fluka) at 30°C for 90 min (HTP130LP; HLC). Acidic protons were subsequently derivatized with 60 μL N-methyl-N-trimethylsilyltri-fluoride (Machery-Nagel GmbH) at 37°C for 30 min. Samples were allowed to cool to room temperature for 10 min. For GC-tof-MS analysis, 60 μL supernatant fluid was transferred into 200-μL glass vials (Chromacol), and 1 μL was injected splitless into a Leco Pegasus III GC-tof-MS system (Leco Inc) consisting of a Focus autosampler (Anatune), an Agilent 6890N gas chromatograph equipped with a DB5-MS column (20 m × 0.25 mm internal diameter × 0.25 μm film). The injector temperature was 250°C, the transfer line was set to 260°C, the ion source temperature was held at 230°C, and the helium flow was 1.2 mL/min. After 1 min at 80°C, the oven temperature was increased by 30°C min to 330°C, held at 330°C for 3 min, and cooled to 80°C. Automated deconvolution and peak finding were performed by using ChromaToF software (Leco). Mass spectra of all detected compounds were compared with in-house standards and spectra in the National Institute of Standards and Technology library (www.nist.gov/srd/nist1.htm) and other publicly available databases. All data pretreatment procedures, including baseline correction, chromatogram alignment, and data compression were performed by using custom scripts in Matlab version 6.5.1 (The Math Works Inc). Targeted peak lists were generated, and peak apex intensities of each characteristic mass in a retention time window were saved in an intensity matrix (run × metabolite).

Data analysis feature selection

Data mining was carried out by following the FIEmspro workflow validated previously in Aberystwyth (27) (http://users.aber.ac.uk/jhd/). PCA was used to reduce data dimensionality and was followed by PC-LDA. Plots of the first 2 DFs allowed visualization of the goodness-of-class separation and the Tw values were used to evaluate the performance of PC-LDA. Discrimination was considered adequate for Tw values >2 and very poor for Tw values <1.0 (27). RF was used in the analysis of the multivariate data, and the RF classification margin was used to assess the classification performance. Average RF margin values ≥0.30 indicate adequate classification in metabolomics experiments (27).

Feature selection techniques were used to select the mass signals responsible for discriminating between different sample classes. A combination of 3 methods—RF, AUC of the ROC, and Welch’s \( t \) test—were used in feature selection to produce a full-feature rank list (27). RF-feature selection was obtained by calculating Importance Scores, ie, the mean decrease in accuracy over all classes when a feature is omitted from the data. The AUC of the sensitivity (true-positive rate) was used against the specificity (false-positive rate), and Welch’s \( t \) test ranked the features by their absolute values of the FDR-corrected \( P \) values.

Randomized resampling strategies with the use of bootstrapping were applied in the process of classification and feature selection to counteract the effect of any unknown, structured variance in the data. In the current data analysis, 100 bootstraps were used for classification and feature selection with RF by using 1000 trees.

Pearson correlation coefficients between selected variables were calculated by using the R-function cor. Variables with correlation coefficients >0.7 were considered to belong to a cluster indicative of different ionization or potential biotransformation/breakdown products of a given food-derived metabolite.

Targeted accurate mass analysis and annotation of FIE-MS signals

Selected nominal mass bins were investigated further by using targeted Nano-Flow (TriVersa NanoMate; Advion Biosciences Ltd) LTQ FT-ICR-MS, where “Ultra” refers to the high-sensitivity ICR-cell). Samples were prepared as for FIE-MS, but at each time point 3 pools of urine samples from randomized groups of participants were prepared and reconstituted in methanol:water (80:20, vol:vol). For each spray, a sample volume of 13.0 μL was used, and 2 μL air was aspirated after the sample. The gas pressure was maintained between 0.2 and 0.6 psi, with the voltage at 1.4–1.7 kV (generally higher for negative ionization mode) and the current at 80–120 nA and −100 and −60 nA in positive and negative ionization modes, respectively. Operating in narrow selective ion monitoring mode, a resolution of 100,000 was chosen and the mass range was scanned for 1 min (50–60 scans). A minimum of 3 biological replicates per class or treatment containing the specific selected mass was required for successful accurate mass verification. The system was calibrated with LTQ-FT calibration solution prepared according to the instrument instructions.

For metabolite signal identification, the accurate mass values were then queried by using MZedDB, an interactive accurate mass annotation tool we developed recently, which can be used directly to annotate signals by means of neutral loss and/or adduct formation rules (28).

FIE-MSn was used for further metabolite signal identification with the scan window set for 20 scans, an isolation width of 1 m/z, and the use of normalized collision energy of 40 V. An activation
coefficient, Q, of 0.250 was chosen, as were an activation time of 30 ms (with wideband activation turned on) and source fragmentation of 20 V. Mass range settings were dependent on the molecular weight of the target ion. Chemical standards investigated with FT-ICR-MS and FIE-MSn were obtained commercially and were of HPLC grade. Standards were prepared by dissolving 1 mg of each metabolite in 1 mL extraction solvent.

RESULTS

Acute exposure to foods of high public health significance is evident in metabolite fingerprints from postprandial urine

FIE-MS and GC-tof-MS metabolite fingerprints were generated for fasting urine samples from 24 MEDE Study 3 participants and for urine samples collected 1.5, 3, and 4.5 h after consumption of the SB (n = 48) and each of the 4 test breakfasts (n = 24) by the same 24 participants. Data were subjected to supervised classification methods to determine whether robust differences in the chemical composition of each urine class were predicted. Of the 3 time points, the 3-h postprandial urine samples provided the strongest classification models (data not shown); thus, all further analysis focused on these samples. In preliminary FIE-MS modeling experiments, the m/z ranges 100–220 and 210–510 in both ionization modes contained signals providing most of the power for discrimination between biological sample classes. Unless stated otherwise, this combined mass range in the ionization mode providing the best classification was used for all further data modeling. PC-LDA score plots of both positive- and negative-ionization modes in FIE-MS and GC-tof-MS data for 3-h postprandial urine samples are shown in Figure 2. We observed good class separation of metabolite patterns detected in positive ionization mode (Figure 2A) in the main vector of discrimination (DF1; Tw value: 3.6) between urine samples collected after salmon consumption from those in urine collected after consumption of the SB and the other test foods. Metabolite patterns in urine collected after the consumption of the breakfasts with broccoli and raspberry components clustered closely together in both ionization modes, but were separated from those for urine collected after exposure to the other foods (Figure 2, A and B). Separation was increased slightly with negative-ionization-mode data (Figure 2B). Metabolite patterns in urine after consumption of corn flakes (standard breakfast) and the whole-grain wheat cereal could not be discriminated from each other in either ionization mode, but were well separated from those in urine after exposure to broccoli, salmon, or raspberry (Figure 2, A and B). Excellent sample class clustering was also achieved by using GC-tof-MS fingerprint data, although it remained difficult to distinguish urine samples collected after consumption of the whole-grain wheat cereal from those after consumption of cornflakes (Figure 2C). Because FIE-MS is higher throughput than GC-tof-MS (18) and, importantly, because signals from FIE-MS can be annotated directly based on subsequent accurate mass measurements (28), we focused on the FIE-MS fingerprinting approach rather than...
on GC-tof-MS for deeper analysis of signals responsible for discrimination between test foods. The magnitude of differences in urine chemistry related to food exposure was assessed between urine samples collected after consumption of each individual test food and of the SB (Table 1). PC-LDA Tw values and RF margin values were used to assess classification performance (27). In general, binary comparisons between each test food and the SB exhibited adequate differences in at least one ionization mode. In particular, urine collected after salmon consumption was strongly discriminated from that collected after each of the other test foods with RF margin values in both ionization modes 0.30 and showed good PC-LDA discrimination (Tw value 2). Urine samples collected after broccoli and raspberry consumption were more difficult to distinguish from urine samples collected after the SB, with RF margin values in positive mode data being just below the 0.30 threshold and poor Tw values. Urinary metabolite patterns after whole-grain wheat cereal consumption were poorly discriminated from equivalent urine samples after the SB in both positive and negative mode data (ie, RF margin values <0.30 and PC-LDA DF1 Tw values <2). Potential biomarkers of whole-grain wheat cereal consumption are not considered further in this analysis.

A population of fasted-state urine samples provides a universal comparator for discovery of biomarker leads indicative of specific dietary exposure

Although the models based on comparison with the SB showed generally adequate performance, the Tw values and RF margins achieved suggest that subsequent data mining to determine the signals responsible for test food discrimination were not optimal (27). Stronger differences in urine metabolome (Tw values >4.0 and RF margins >0.6) were found when the urinary metabolome after exposure to each test breakfast was compared with the metabolome of a population of urine samples collected in the fasting state before exposure to the SB (Table 2). In this comparison, the fasting urine samples (n = 48) were collected from

TABLE 2
Data models derived from flow infusion electrospray–ionization mass spectrometry fingerprinting to discriminate between metabolite patterns in 48 fasting urine samples and those in urine collected 3 h after consumption of the SB and of test breakfasts in which the cornflakes and milk component of the standardized breakfast was replaced with each of 4 different test foods

<table>
<thead>
<tr>
<th>Test food</th>
<th>Positive-ion fingerprints</th>
<th>Negative-ion fingerprints</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RF classification margin</td>
<td>DF1&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Tw</td>
<td></td>
</tr>
<tr>
<td>SB</td>
<td>0.72</td>
<td>8.69</td>
</tr>
<tr>
<td>Salmon</td>
<td>0.76</td>
<td>7.92</td>
</tr>
<tr>
<td>Broccoli</td>
<td>0.70</td>
<td>6.46</td>
</tr>
<tr>
<td>Raspberry</td>
<td>0.71</td>
<td>7.76</td>
</tr>
<tr>
<td>Whole-grain wheat cereal</td>
<td>0.70</td>
<td>6.98</td>
</tr>
</tbody>
</table>

<sup>1</sup> Pairwise RF classification and principal components–linear discriminant analysis of data acquired by flow infusion electrospray–ionization mass spectrometry (100–510 m/z; positive- and negative-ion modes) in post–test breakfast urine samples (salmon, broccoli, raspberry, and whole-grain wheat cereal; n = 24) compared with the standardized breakfast (cornflakes; n = 48) from 24 individuals; data were log<sub>10</sub> transformed and normalized by using the sample total ion count. DF, discriminant function; RF, Random Forest; Tw, eigenvalue.

<sup>2</sup> Tw, DF1 eigenvalue.
the MEDE Study 3 participants \((n = 24)\) before consumption of the SB. It was considered that this population of fasting urine samples would provide a representative, standardized comparator for discovery of biomarker leads indicative of specific dietary exposure. The postprandial urine samples were also compared against the fasting urine samples collected on the same day in MEDE Study 3 \((n = 24)\), all the fasting urine samples in MEDE Study 2 \((n = 6 \times 24)\), and fasting urine samples collected in MEDE Study 2 \((n = 24)\) \((20)\). Strong models discriminating between fasting and postprandial urine samples were achieved in each case, with \(T_w\) values \(>4.0\) and RF margins \(\geq 0.6\) (data not shown).

A combination of modeling statistics from 3 feature selection methods (RF, AUC, and Welch’s \(t\) test), together with the FDR-adjusted \(P\) values, were used to rank metabolite signals \((m/z)\) for discrimination power (see Supplemental data 3 and 4 under “Supplemental data” in the online issue for full positive- and negative-ionization-mode data, respectively). To avoid potential random effects, resampling was done with the use of the bootstrap method. With the use of the appropriate significance thresholds (RF importance score threshold of \(>0.002\), FDR-adjusted \(P\) value of \(<1 \times 10^{-5}\) and AUC \(>0.90\), respectively), \(~15–30~\) FIE-MS metabolite signals \((m/z)\) provided most of the classification power to discriminate fasting-state urine samples from urine samples collected 3 h after consumption of each test food. Detailed examination of the top ranked signals in each binary comparison showed considerable overlap (see Supplemental data 3 and 4 under “Supplemental data” in the online issue), which is perhaps unsurprising because all test meals include the “core” components of the SB, ie, orange juice, tea with skimmed milk and sugar, and butter croissant.

**Data filtering allows identification of metabolite signals associated specifically with exposure to test foods**

Postprandial urine samples collected at 1.5, 2, 3, 4, and 4.5 h after consumption of the SB were available from either the MEDE Study 2 or MEDE Study 3 phases of the MEDE project. RF analysis had been used previously \((20)\) to rank \(m/z\) signals discriminating fasting from postprandial urines at all 5 time points. Further analysis of these data identified 36 metabolite signals with an RF importance score \(>0.002\) in \(\geq 2\) of the 5 time points, shown in ascending order of mass in column 1 of Table 3. Using salmon as an example test food, we observed that signals associated with SB consumption are also highly ranked for explanatory power and are interspersed with further unique signals (footnote 6 in column 3, Table 3) that are presumably related to salmon consumption. To highlight the metabolite signals associated uniquely with salmon exposure, these positive ion signals present in urine after the SB but absent from fasting urine were filtered from the data matrices and the feature selection repeated (see Supplemental data 5 under “Supplemental data” in the online issue). After this data filtering, the signals expected to be more specific for salmon exposure were then ranked highly (footnote 6 in column 6, Table 3) and had RF importance scores \(>0.002\). RF importance score analysis, in which urine samples collected after consumption of each test food were compared with fasted-state urine after data filtering, showed a smaller number of positive ion signals with explanatory power (see Supplemental data 5 under “Supplemental data” in the online issue). The same filtering method was used to remove the negative ionization mode nominal mass signals, which discriminated the SB from fasting urine from the broccoli and raspberry data matrices, and the feature selection process was repeated (see Supplemental data 6 under “Supplemental data” in the online issue). Examination of the top highly ranked signals showed that many of the explanatory signals were unique to each test food but, in the case of raspberry and broccoli, there was some overlap at the lower rankings (see Supplemental data 6 under “Supplemental data” in the online issue).

Approximately half of the potentially explanatory signals exhibited an increase in intensity in postprandial urine samples in relation to the fasting state (indicated by the highlighted red “U” in Supplemental data 5 and 6 under “Supplemental data” in the online issue). Such signal behavior would be desirable for development of biomarkers of dietary exposure, and so these nominal mass bins were investigated further. The heat map (Figure 3) shows the relative signal intensity of these targeted masses in urine samples collected 3 h after consumption of each test food compared with fasting. Almost all of the most intense signals (shaded deep red) selected in urine collected after exposure to salmon were elevated specifically, and strongly, by this acute food challenge. The strongest discriminatory signals identified after raspberry consumption were elevated specifically in response to this fruit; however, there was overlap with signals from broccoli (metabolite signals \(m/z 211\) and \(m/z 176\)). Some of the explanatory signals elevated with broccoli exposure were also observed after eating raspberries and salmon but were present at a lower relative intensity level.

**Identification of explanatory metabolite signals indicates strong links to acute dietary exposure**

Accurate mass analysis by FT-ICR-MS was carried out in all explanatory nominal mass bins suspected to have value as potential biomarkers of salmon, raspberry, and broccoli consumption (Figure 3). Most bins targeted had a single dominant ion, and only 2 bins could not be monitored because of insufficient signal for the less sensitive FT-ICR-MS analytic procedure (Table 4). Accurate mass information was used to interrogate MZedDB \((28)\), at 1 ppm resolution; in most instances, a single identity was suggested (Table 4). Further fragmentation analysis (FIE-MSn) including the use of pure chemical standards, comparisons with spectra in publicly available data depositories and literature \((29, 30)\) (see Supplemental data 7 under “Supplemental data” in the online issue) confirmed the identity of most of the targeted signals representing exposure to the 3 test foods.

The results of \(m/z\) annotation suggested that many signals were related to the consumption of metabolites derived from each of the test foods (Figure 4). Exposure to smoked salmon was associated with the appearance in urine of 1-methylhistidine and anserine (\(\beta\)-alanyl-1-methylhistidine), which are metabolites of histidine and histidine-derived dipeptides (Figure 4A) and are known to be abundant in the skeletal muscle of many oily fish species, including salmon \((22, 31)\). In addition, consumption of smoked salmon was followed by increased urinary concentrations of TMAO, a metabolite of trimethylamine (Figure 4B) —a degradation product formed from carnitine \((32)\) and found at high concentrations in fish \((33, 34)\). Similarly, exposure to both broccoli and raspberry resulted in increases in ascorbate which
is abundant in many fruits and vegetables (35, 36). After broccoli consumption (Figure 4C), putative ascorbate catabolism products (37) including tetronic acids and their possible derivatives, together with 5-carbon compounds including xylonate/lyxonate (or possible stereoisomers) were also detected. Raspberry consumption resulted in the appearance in urine of products (37) including tetronic acids and their possible derivatives, together with 5-carbon compounds. Compounds derived from the major polyphenols found in orange juice (naringenin glucuronide and hesperetin glucuronide) were also selected as the top 20, but lower ranking, negative-ion explanatory signals detectable in 3-h postprandial urine after consumption of the test breakfasts containing either broccoli or raspberry (Table 4) or salmon (see Supplemental data 6 under “Supplemental data” in the online issue), but not the whole-grain wheat cereal (see Supplemental data 6 under “Supplemental data” in the online issue).

DISCUSSION
The development of biomarkers of intake of specific foods that can be measured in readily accessible body fluids is difficult. The challenge is to discover, identify, and then to validate chemicals...
The use of acute phase postprandial urine samples for biomarker lead discovery is complicated by the dynamic shifts in metabolism after the transition from the fasted to the fed state (39) and during the intermeal periods, coupled with inherent (40) and substantial, inter- and intraindividual variability in human metabolite profiles (16, 40, 41). Despite these major sources of variability, we showed that nontargeted metabolite fingerprinting of acute phase postprandial urine samples by using MS, followed by accurate mass analysis, can identify metabolite signals with potential as biomarkers for exposure to specific foods of high public health significance. The key to success of the suggested biomarker discovery strategy was a data-filtering approach based on the masking of a group of dominant signals in FIE-MS fingerprints of urine samples found to be present in urine collected 2–4 h after consumption of an SB. It is pertinent to note that this high-throughput, hierarchal analytic approach avoids any requirement for extensive preprocessing of MS chromatographic data (15, 18) and, in general, achieves a very high success rate for de novo metabolite annotation (based on ultrahigh accurate mass) in comparison with methods using NMR or GC-MS (15–21, 24, 33, 41). In addition, the current data suggest that fasted-state urine samples can provide a universal comparator for the discovery of biomarker leads indicative of specific dietary exposure.

A reassuring aspect of the current results was that many of the signals identified as potential dietary biomarkers of recent exposure to the 3 test foods could be linked to metabolites or bio- transformation products of metabolites reported to be present in each test food (22, 31, 33, 35, 36). For example, TMAO, anserine, and 1-methylhistidine were present at greatly elevated concentrations in urine after consumption of smoked salmon. Other researchers have reported elevated urinary TMAO concentrations after the consumption of seafood (16), and our current analysis showed the appearance of TMAO in urine after salmon consumption. The concurrent appearance of elevated concentrations of 1-methylhistidine likely resulted from the metabolism of ingested anserine (22), which is found at relatively high concentrations in the muscle of many oily fish species (22, 31). Anserine is found at lower concentrations in chicken and turkey breast meat along with carnosine—a related histidine-derived dipeptide (42, 43). In contrast, red meats contain relatively low concentrations of anserine, but have relatively high concentrations of carnosine (43, 44). Although 1-methylhistidine may not provide a biomarker specifically for dietary exposure to oily fish, its concentration in urine is unlikely to reflect the general level of longer-term (mammalian) meat intake as suggested previously (43). Further studies using a multiplex analysis of the relative concentrations in urine of several muscle-derived signals, including carnosine, anserine, 1-methylhistidine and TMAO, may provide useful information on the relative intakes of fish and poultry compared with red meat.

Our observation that the consumption of both broccoli and raspberry resulted in an increase in the postprandial urine concentrations of ascorbate and its suspected breakdown products is expected given the fact that such fruit and vegetables are rich sources of vitamin C. This agrees with reports that plasma ascorbate concentrations reflect fruit and vegetable intake (33, 34). The measurement of these novel ascorbate metabolites in urine could provide a more robust alternative to the quantification of plasma vitamin C concentrations, because such methods have problems with reproducibility (45). Both FIE-MS fingerprinting and GC-tof-MS profiling of postprandial urine suggest that reflecting exposure to specific dietary components against a highly variable background of signals that may be common to many foods. Overall, our results showed that a well-validated protocol for handling and sample volunteers (20), in combination with state-of-the-art metabolite fingerprinting (18) and data-mining (27) procedures has great potential for the discovery of dietary exposure biomarker leads. The future application of any biomarker leads for measuring habitual dietary intake will require further work, particularly to investigate their specificity and sensitivity to report exposure to specific food or food groupings in an epidemiologic context. Encouragingly, our own recent work (21) and previous (25, 26) observations in relation to citrus fruit exposure showed that, although the excretion rate of proline betaine peaked 2–4 h after consumption of orange juice, it may be a good indicator of habitual citrus exposure as well as acute intake.

**FIGURE 3.** Changes in signal intensity in urinary potential biomarkers after consumption of each of 3 test foods and of the SB. Heat map representation of FIE-MS data generated by the analysis of 3-h postprandial urine samples, in which the more intense red color denotes an increase in signal intensity compared with fasting urine [after a 12-h (minimum) fast], and the more intense green color denotes a decrease in signal intensity compared with fasting urine. Yellow indicates no change compared with fasting. Key top-ranked potential biomarkers associated with the consumption of smoked salmon [A; positive-mode (+ve) FIE-MS data], broccoli [B; negative-mode (+ve) FIE-MS data] and raspberry [C; negative-mode FIE-MS data] compared with fasting urine. FIE-MS, flow infusion electrospray–ionization mass spectrometry; SB, standardized breakfast.
The utility of the potential dietary exposure biomarkers identified by using FIE-MS fingerprinting in the current study could be evaluated in existing epidemiologic data based on metabolomic analysis of 24-h urine samples, as described recently (25). Similarly, biomarker potential could be evaluated by de novo targeted analysis of either 24-h or “spot” urine samples derived from recent nutritional intervention studies (7, 15, 33, 41, 29). In addition to nutritional studies, the specific signals identified in the current study are putative biomarkers of specific foods, and these may have wider utility in studies of food exposure, in which the interest may be in nonnutritional components of the foods (eg, those present naturally or as a consequence of food processing or food contamination).

In summary, in the MEDE Study, we developed and tested a biomarker lead discovery strategy that can identify urinary metabolites appearing in urine after exposure to specific foods. We observed that 1-methylhistidine is strongly associated with salmon consumption and may have value in discriminating between types of meat and fish in diets. Measurement of the novel ascorbate metabolites discovered in urine after consumption of broccoli and raspberries could provide a more robust alternative to the quantification of vitamin C concentrations in plasma as a marker of general fruit and vegetable intake. Similarly, elevated concentrations of sulphonated derivatives of cafffeic acid and methyl-epicatechin in urine may reflect more specific exposure to certain subgroupings of fruit, such as raspberries.

Table 4: Identity of metabolites in explanatory mass bins associated with the consumption of specific foods consumed in test breakfasts obtained by using accurate mass analysis and tandem mass spectrometry

<table>
<thead>
<tr>
<th>Potential mass</th>
<th>Major signal in FT-ICR-MS</th>
<th>Putative metabolite (1 ppm)</th>
<th>Ionization product</th>
</tr>
</thead>
<tbody>
<tr>
<td>241</td>
<td>241.12964 Anserine</td>
<td>[M+H]^+</td>
<td>13C isotope [M+H]^+</td>
</tr>
<tr>
<td>242</td>
<td>242.1302 Anserine</td>
<td>[M+H]^+</td>
<td>13C isotope [M+H]^+</td>
</tr>
<tr>
<td>192</td>
<td>192.07427 1- and 3-Methylhistidine</td>
<td>[M+Na]^+</td>
<td></td>
</tr>
<tr>
<td>208</td>
<td>208.04825 1- and 3-Methylhistidine</td>
<td>[M+K]^+</td>
<td></td>
</tr>
<tr>
<td>279</td>
<td>279.08567 Anserine</td>
<td>[M+K]^+</td>
<td></td>
</tr>
<tr>
<td>221</td>
<td>221.06445 Unknown</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>151</td>
<td>151.14409 Trimesylamine-N-oxide</td>
<td>[2M+H]^+</td>
<td></td>
</tr>
<tr>
<td>259</td>
<td>259.02808 Caffeoyl sulfate</td>
<td>[M–H]^-</td>
<td></td>
</tr>
<tr>
<td>211</td>
<td>211.00160 Ascorbate</td>
<td>[M+Cl]^+</td>
<td></td>
</tr>
<tr>
<td>383</td>
<td>383.04419 Methyl-epicatechin sulfate</td>
<td>[M–H]^-</td>
<td></td>
</tr>
<tr>
<td>341</td>
<td>341.12405 Unknown glucuronide</td>
<td>[M–H]^-</td>
<td></td>
</tr>
<tr>
<td>194</td>
<td>194.04603 3-Hydroxyhippuric acid</td>
<td>[M–H]^-</td>
<td></td>
</tr>
<tr>
<td>447</td>
<td>447.09330 Naringenin glucuronide</td>
<td>[M–H]^-</td>
<td></td>
</tr>
<tr>
<td>211</td>
<td>211.00160 Ascorbate</td>
<td>[M+Cl]^+</td>
<td></td>
</tr>
<tr>
<td>135</td>
<td>135.02996 Tetronic acids</td>
<td>[M–H]^-</td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>103.04010 Tetronic acid derivative</td>
<td>[M–H]^-</td>
<td></td>
</tr>
<tr>
<td>165</td>
<td>165.04051 t-Xylonate-L-lyxonate</td>
<td>[M–H]^-</td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>104.04344 Tetronic acid derivative</td>
<td>13C isotope [M–H]^–</td>
<td></td>
</tr>
<tr>
<td>121</td>
<td>121.05068 Threitol/erythritol</td>
<td>[M–H]^-</td>
<td></td>
</tr>
<tr>
<td>447</td>
<td>447.09330 Naringenin glucuronide</td>
<td>[M–H]^-</td>
<td></td>
</tr>
<tr>
<td>477</td>
<td>477.10421 Naringenin glucuronide</td>
<td>[M–H]^-</td>
<td></td>
</tr>
</tbody>
</table>

Accurate mass analysis by using FT-ICR-MS and FIE-MSn with pure chemical standards and comparisons with spectra from publicly available data depositories (see Supplemental Table 7 under “Supplemental data” in the online issue). FIE-MSn, flow infusion electrospray–ionization tandem mass spectrometry; FT-ICR-MS, Fourier transform–ion cyclotron resonance ultra-mass spectroscopy; ppm, parts per million.
multiplex analysis of potential biomarkers in the future may be helpful in the investigation of links between habitual diet and health outcomes.

We thank the volunteers for their commitment; the Clinical Research Facility Centre for nursing support; Claire Kent, Heather E Gifford, Julie Coaker, and Linda Penn for their practical support; and Marks & Spencer for donating the chocolate éclairs used in the standardized evening meal.

The authors’ responsibilities were as follows—JD, JCM, and MEB: designed the research; GF: recruited the participants and collected the metadata; GF: supervised the dietary interventions and sampling; LX: provided technical support during dietary interventions and sampling; GF, MEB, and WL: refined the sampling method; MEB: designed and supervised the metabolite fingerprinting experiments and preprocessed the data for analysis; KT: performed the FIE-MS and GC-tof-MS analysis; AJL: performed the statistical analysis of metabolite fingerprint data and the targeted accurate mass analysis of explanatory signals; AJL and JD: performed the signal annotations, interpreted the data, and produced the figures; JD, AJL, JCM, MEB, and GF: wrote the manuscript; and JD: had primary responsibility for the final content. All authors read and approved the final manuscript. None of the authors had a conflict of interest.

REFERENCES


