Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption1–3
Silke S Heinzmann, Ian J Brown, Queenie Chan, Magda Bictash, Marc-Emmanuel Dumas, Sunil Kochhar, Jeremiah Stamler, Elaine Holmes, Paul Elliott, and Jeremy K Nicholson

ABSTRACT
Background: New food biomarkers are needed to objectively evaluate the effect of diet on health and to check adherence to dietary recommendations and healthy eating patterns.
Objective: We developed a strategy for food biomarker discovery, which combined nutritional intervention with metabolic phenotyping and biomarker validation in a large-scale epidemiologic study.
Design: We administered a standardized diet to 8 individuals and established a putative urinary biomarker of fruit consumption by using 1H nuclear magnetic resonance (NMR) spectroscopic profiling. The origin of the biomarker was confirmed by using targeted NMR spectroscopy of various fruit. Excretion kinetics of the biomarker were measured. The biomarker was validated by using urinary NMR spectra from UK participants of the INTERMAP (International Collaborative Study of Macronutrients, Micronutrients, and Blood Pressure) (n = 499) in which citrus consumption was ascertained from four 24-h dietary recalls per person. Finally, dietary patterns of citrus consumers (n = 787) and nonconsumers (n = 1211) were compared.

Results: We identified proline betaine as a putative biomarker of citrus consumption. High concentrations were observed only in citrus fruit. Most proline betaine was excreted < 14 h after a first-order excretion profile. Biomarker validation in the epidemiologic data showed a sensitivity of 86.3% for elevated proline betaine excretion in participants who reported citrus consumption and a specificity of 90.6% (P < 0.0001). In comparison with noncitrus consumers, citrus consumers had lower intakes in fats, lower urinary sodium-potassium ratios, and higher intakes of vegetable protein, fiber, and most micronutrients.

Conclusion: The biomarker identification and validation strategy has the potential to identify biomarkers for healthier eating patterns associated with a reduced risk of major chronic diseases. The trials were registered at clinicaltrials.gov as NCT01102049 and NCT01102062.

INTRODUCTION
Nutritional factors play a major underlying role in the causation of the global burden of chronic disease; specifically, a healthy diet rich in fruit and vegetables is associated with lower rates of cancers, diabetes, cardiovascular diseases, and related risk factors such as raised blood pressure and serum cholesterol (1–6). Focus has shifted from examining single nutrient relations with disease towards analyzing complex nutrient interactions and dietary patterns to define a more holistic relation between nutrition and associated diseases (7, 8). A food pattern high in fruit, vegetables, fish, whole grains, and legumes shows inverse correlations with features of metabolic syndrome (9), risk of colorectal cancer (10), and adverse blood pressure and serum lipid profiles (5, 11, 12).

Methods to assess dietary intakes of free-living populations rely mainly on questionnaire data that are subject to possible reporting and other biases (13). Objective measures that use biomarkers are needed to validate dietary assessment and check adherence to dietary recommendations and healthy eating patterns, but few such biomarkers are available (14), including markers for fruit and vegetable intake (15). The development of robust food biomarkers may help to improve disease risk stratification by better characterizing the metabolic phenotype at the individual level. Numerous studies that addressed single food components or nutrients have been conducted in small scale laboratory studies but few, if any, have been translated into free-living population studies. We show the use of high-throughput screening by 1H nuclear magnetic resonance (NMR) spectroscopy for citrus fruit biomarker discovery and its application to a large-scale population survey. High-resolution spectral analyses, typically NMR or mass spectrometry, have been used to

1 From Biomolecular Medicine, Department of Surgery and Cancer, Faculty of Medicine (SSH, M-ED, EH, and JKN), and the Department of Epidemiology and Biostatistics (IJB, QC, MB, and PE), Imperial College London, London, United Kingdom (EH, PE, and JKN).
2 Supported by Nestlé Nectec (to SSH). The International Collaborative Study of Macronutrients, Micronutrients, and Blood Pressure (INTERMAP) is supported by grants RO1 HL50490 and RO1 HL084228 from the National Heart, Lung, and Blood Institute, Bethesda, MD. The UK samples were funded by a project grant from the West Midlands National Health Service Research and Development and grant number R2019EPH from the Chest, Heart and Stroke Association, Northern Ireland.
3 Address correspondence to JK Nicholson, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, South Kensington, SW7 2AZ, London, United Kingdom. E-mail: j.nicholson@imperial.ac.uk; or to P Elliott, Department of Epidemiology and Biostatistics, Imperial College London, St Mary’s Campus, London W2 1PG, United Kingdom. E-mail: p.elliott@imperial.ac.uk.

Received April 13, 2010. Accepted for publication May 26, 2010.
generate metabolic signatures from biological samples and ob-
tain complex profiles of a wide range of metabolite classes (16–
18). Population-based studies have shown marked differences in
metabolic profiles within and between populations that reflect,
in part, dietary differences as key components of the complex
interplay between environmental and genetic influences on
disease risk (19); effects of dietary interventions on the met-
abolic phenotype have also been explored (20–23). In the current
study, we outline a strategy for a biomarker discovery for healthy
eating that is exemplified by citrus fruit consumption. Specifically,
we combined nutritional intervention, metabolic profiling and
biomarker cross-validation in large-scale epidemiologic data.

SUBJECTS AND METHODS

Fruit-intervention study

A laboratory study was designed to detect urinary biomarkers of
fruit consumption by using a nontargeted metabolic profiling
approach. The study was a fruit-meal intervention that involved
8 volunteers (7 women and 1 man; age range: 28–45 y) who met
the following inclusion criteria: participants were healthy, aged
18–45 y, and nonsmokers and had a body mass index (BMI; in
kg/m²) 18–25, an absence of regular drug intake and regular
food supplement intake, and no antibiotic use within the pre-
vious 3 mo. Participants consumed a standardized breakfast,
lunch, and dinner that comprised whole-grain bread and cheese
(breakfast), a ham sandwich (lunch), and pasta and tomato sauce
(dinner) from the run-in day (day 0) until lunch on day 3. In
addition to the standard dinner, a supplementary mixed-fruit
meal (apple, orange, grapes, and grapefruit) was introduced on
the evening of day 2. Urine was collected 4 times/d (first
morning urine, before lunch, before dinner, and at bed time),
from the morning of day 1 until the evening of day 3. Urine
specimens were collected into sterile tubes (Sterilin, Aberbargoed,
United Kingdom) and stored at −40°C until analysis. The study
was approved by Imperial College London Research Ethics
Committee.

Food-analysis study

To ascertain the relative concentrations of proline betaine in
fruit and fruit juices and to obtain an estimate of the specificity of
proline betaine as a marker of citrus consumption, the relative
concentrations of proline betaine were assessed in a number of
fruit and commercially available fruit juices. Proline betaine
quantification of fruit juices and fruit extractions was carried out
with a fully relaxed water-presaturated 1H NMR technique.
Three samples for each citrus fruit and citrus fruit juice and one
sample for other fruit and fruit juices were analyzed. For the
fruit extracts, whole fruit was homogenized, centrifuged for
5 min at 16,000 × g, and the supernatant fluid was used as a fruit
extract. An aliquot of 200 μL fruit juice or fruit extract was
dissolved in 800 μL phosphate D₂O buffer (0.1 mol/L; pH 7.4)
that contained 0.35 mmol sodium 3-(trimethylsilyl)-propionate-
d₄ (TSP). Samples were vortexed and centrifuged for 5 min at
16,000 × g. A 600-μL sample of the supernatant fluid was
transferred into 5-mm outer-diameter NMR tubes for analysis by
1H NMR spectroscopy.

Assessment of the excretion profile of proline betaine after
orange juice consumption

A study was undertaken to characterize the excretion profile of
proline betaine after orange juice consumption. Six volunteers
(4 women and 2 men; age range: 24–36 y) who met the inclusion
criteria of being healthy, aged 18–45 y, and nonsmokers and had a
BMI of 18–25 kg/m² and an absence of regular drug intake and
regular food supplement intake were recruited. Participants
consumed a restricted diet that excluded all citrus fruit, ethnic
foods, alcohol, grain legumes, and cheese. Day 0 was a run-in
day; spot urine specimens were collected on day 1 (ie, baseline
samples, 6 times/d, collected at 0800, 1000, 1200, 1600, and
2000 and at bed time); an orange juice challenge (250 mL or-
geJuice) was administered on day 2 at 1000, and spot urine
was collected at the same times as on day 1 and continued until
1000 on day 3 (study finish). Urine specimens were collected in
sterile tubes (Sterilin) and stored at −40°C until analysis. The
study was approved by Imperial College London Research Ethics
Committee.

1H NMR spectroscopic analyses

1H NMR spectra were acquired on a Bruker 600 MHz spec-
trometer (Bruker Analytische GmbH, Rheinstetten, Germany)
that operated at 600.13 MHz. The spectrometer used a standard
one-dimensional pulse-sequence [recycle delay (RD)-90°-1-
90°-mixing time (tm)-90°-acquire] free-induction decay (FID)
with water-suppression irradiation during a RD of 2 s with tm set
to 100 ms and a 90° pulse set to 10 μs. Spectra were acquired
with 128 scans into 32-K data points with a spectral width of
12,000 Hz. Fruit spectra were acquired by using a standard one-
dimensional pulse sequence [RD-90°-τ-90°-tm-90°-acquire
FID] with water suppression that irradiated during a RD of 2 s with
tm set to 100 ms and a 90° pulse set to 10 μs. The proline
betaine and TSP resonances were fully relaxed after a delay time
tau of 1.25 and 3 s, respectively, resulting in a spin-lattice re-
 laxation time T1 (=tau/ln2) of 4.33 s. The interpulse delay time
d1 (=5 × T1) was therefore set to 21.64 s (24). For spectral
processing, FIDs were multiplied by an exponential function
that corresponded to line broadening of 0.3 Hz before Fourier
transformation. All spectra were manually phased, baseline
corrected, and calibrated to TSP [δ (chemical shift) 0] and
exported into Matlab software (2009a; The MathWorks Inc,
Natick, MA) for data analysis. The spectra were normalized by
using a probabilistic quotient normalization algorithm (25) to
account for urinary dilution effects and aligned by using a re-
cursive segment-wise peak-alignment algorithm (26).

Multivariate data analyses

All timed 24-h urine specimens of the nutrition study that
were collected on day 1 in the morning until evening of day 2
(before the fruit meal) were defined as standard class; specimens
collected after the fruit meal (bed time of day 2 until the evening
of day 3) were defined as fruit class. Pairwise partial least-
squares discriminant analysis (PLS-DA) (27) was carried out
with Matlab software (2009a; The MathWorks Inc). The PLS-
DA loading plots were created with an in-house Matlab (2009a;
The MathWorks Inc) script according to the method of Cloarec
et al (28).
Evaluation of urinary proline betaine excretion in an epidemiologic study and comparison with dietary recall data

The use of proline betaine excretion as a marker of citrus intake and as a surrogate marker of healthier eating patterns was evaluated by using a multipopulation epidemiologic study (International Collaborative Study of Macronutrients, Micronutrients, and Blood Pressure (INTERMAP)) with robust dietary recall questionnaire data. INTERMAP is a large-scale epidemiologic study that involves men and women aged 40–59 y from 17 population samples in China, Japan, the United Kingdom, and the United States. The study investigates the role of multiple dietary factors in causes of raised blood pressure, which includes the use of the metabolome-wide association approach to identify blood pressure–related biomarkers of dietary patterns (19). Individuals were selected randomly from population lists and stratified for age and sex. A 24-h urine specimen and 2 consecutive in-depth multipass 24-h dietary recalls were collected from each participant, and the procedure was repeated after 3 wk (29, 30). The two 24-h urine collections were obtained on days 2 and 4, which corresponded to the first and second and third and fourth dietary recalls, respectively. For the prediction of citrus fruit intake on the basis of proline betaine excretion, we used the Belfast (UK) sample (n = 220) as the training set and the West Bromwich (UK) sample (n = 279) as the validation set. We quantified proline betaine by peak integration (δ 3.106–3.116) by using the 24-h urinary 1H NMR spectral data from the first of the two 24-h collections from each participant. Standard sample-preparation procedures, 1H NMR–acquisition variables, and spectral processing methods were used as described by Holmes et al (19). INTERMAP received institutional ethics committee approval for each site, all participants gave written consent, and all procedures were in accordance with institutional guidelines. The 24-h dietary recall data for the 2 UK-population samples (Belfast and West Bromwich) were coded by using the UK FOODBASE database (version 1.3; Institute of Brain Chemistry, London, United Kingdom) (31). From this database, information on all citrus fruit and citrus fruit juices consumed was obtained. Citrus consumption was classified as any case in which citrus fruit had been reported as being consumed, regardless of the quantity, with the exception of lemon or lime juice used in cooking because the amount used in such cases generally was <10 g.

Receiver operating characteristic–curve comparison of 1H NMR calculated urinary proline betaine excretion values with 24-h food-record data

A receiver operating characteristic (ROC) curve was constructed by using information on citrus fruit consumption in the 24-h food-record data for identification of citrus consumers and noncitrus consumers. Every participant who recorded citrus consumption was defined as a citrus consumer. The ROC curve was a plot of sensitivity compared with 1 – specificity for all possible values of cutoffs from proline betaine integrals of the NMR spectral data.

The optimal operating point [shortest distance from the optimal point (0,1) to the intersect of the ROC curve (32)] was used to measure the optimal cutoff for which specificity and sensitivity were calculated.

FIGURE 1. Identification of putative biomarkers by using metabolite profiling and multivariate analysis. A: Study design for the dietary intervention study (n = 8). B: Representative 1H nuclear magnetic resonance (NMR) spectra of urine specimens in response to fruit consumption (red) compared with the standard (STD) meal (black). Apparent differences are highlighted (dashed rectangles). C: Partial least-squares discriminant analysis (PLS-DA) scores plot of urine specimens 0–24 h after fruit challenge, which shows a clear separation of the fruit and STD meals. All urine specimens from the morning of day 1 to the evening of day 2 were allocated to the STD diet, and all urine specimens collected after consumption of the fruit meal (bed time of day 2 until evening of day 3) were allocated to the fruit class. D: Loading plots of the fruit challenge compared with the STD meal indicated the following putative biomarkers for fruit consumption: hippuric acid (δ 2.97d, 7.55t, 7.64t, and 7.84d), proline betaine (δ 2.18m, δ 2.30m, δ 2.50m, δ 3.11, δ 3.31, and δ 3.54), tartaric acid (δ 4.34s), and unknown (δ 7.74d and δ 6.98d). The P value of proline betaine before fruit consumption compared with after fruit consumption was <0.0001. ppm, parts per million; a.u., arbitrary units; d, doublet; t, triplet; m, multiplet; Y, response variable (classification identifier); R2Y, variation of Y modeled; Q2Y, cross-validated variation of Y predicted; T[1], first predictive PLS scores vector; Tyosc [1], first orthogonal PLS score vector.
Classification of citrus consumption on the basis of urinary proline betaine concentrations and association of spectroscopically calculated citrus consumption with nutrient intakes, BMI, and blood pressure

Participants from the INTERMAP US and UK samples were classified as citrus consumers or noncitrus consumers by using the urinary proline betaine integral. Individuals with proline betaine >39.4 in both 24-h urine collections (n = 787) were classified as citrus consumers, and noncitrus consumers were defined as individuals with a urinary proline betaine integral \( \leq 39.4 \) in both 24-h collections (n = 1211). Individuals with a urinary proline betaine integral >39.4 in one 24-h collection only (n = 645) were excluded. Mean values of selected nonnutrient and nutrient variables from all four 24-h dietary recalls were compared between the 2 groups with adjustment for age, sex, and country with SAS 9.1 software (SAS Institute, Cary, NC).

RESULTS

Untargeted \(^1\)H NMR metabolite profiling

To identify novel biomarkers of fruit consumption, we conducted a human dietary intervention trial (n = 8) and evaluated the effect of a mixed-fruit meal, which consisted of apples, grapes, oranges, and grapefruit, on the urinary metabolome (Figure 1A). Participants’ diets were standardized over 4 d, and the mixed-fruit meal was introduced on day 2. Urine was collected 4 times/d for 3 d and analyzed by using multivariate modeling of spectroscopic profile data. The urinary metabolite profiles, as measured by 600 MHz \(^1\)H NMR spectroscopy, revealed apparent metabolic changes in urine profiles between prefruit and postfruit meal consumption (Figure 1B). Further biochemical information was extracted via application of PLS-DA (Figure 1, C and D) in which we observed an increased excretion of proline betaine, tartaric acid, hippuric acid, and an unknown benzoic acid metabolite after fruit consumption compared with after consumption of a baseline diet (Figure 1D). Of those compounds, proline betaine showed the strongest correlation (\( r^2 = 0.4; P < 0.0001 \)) with fruit ingestion; therefore, we investigated its dietary origin and characterized its urinary excretion kinetics.

Quantification of proline betaine in fruit and fruit juices

To assign the origin of urinary proline betaine excretion after the fruit challenge, we measured concentrations of proline betaine in selected fruit and commercially available fruit juices by using a standard \(^1\)H NMR experiment optimized for quantification of this compound (Table 1). All citrus fruit tested contained proline betaine. Concentrations of proline betaine varied depending on the type of citrus fruit (orange > lime > satsumas > grapefruit > lemon) and the method of juice processing (commercially processed > freshly squeezed) with values \( \leq 1316 \) mg proline betaine/L for orange juice from concentrate. In contrast, the proline betaine signal was of a low intensity in spectra of the other commonly available fruit and fruit juices tested (Table 1).

Excretion kinetics of proline betaine after orange juice consumption

We investigated proline betaine as a potential biomarker for citrus fruit intake by measuring urinary proline betaine excretion in 6 individuals after consumption of 250 mL orange juice. Urine specimens were collected before (-26 to 0 h) and after (+2 to +24 h) the orange juice challenge (Figure 2, A and B). The singlet peak at \( \delta 3.11 \), which represented the CH3 moiety of proline betaine, showed a minimal overlap with other peaks in the spectrum. Excretion of proline betaine was rapid and peaked in all individuals at the 2-h postintervention collection, as calculated from the spectral integral [spectra normalized to probabilistic quotient (25) over the region \( \delta 3.106–3.116 \)]. In all participants, concentrations declined to almost baseline after >24 h, with most proline betaine excretion occurring in the first 14 h (83% of 24-h excretion) (Figure 2B).

Validation in epidemiologic data of proline betaine as a biomarker for citrus fruit intake

To validate our experimental findings, we investigated proline betaine as a potential biomarker of citrus fruit consumption in a free-living human population by using \(^1\)H NMR spectral data obtained from the 2 INTERMAP UK urine collections per individual (29). The technical error of proline betaine quantification from \(^1\)H NMR spectra was 2.64% calculated from data on specimens split in the field and blinded to the laboratory (n = 40; 8% of total samples) by using the formula

\[
\left[ \frac{(\Sigma d^2 + 2N)^{0.5} \times 100}{\bar{x}} \right]
\]

where \( d \) is the within-pair difference, \( N \) is the number of split-sample pairs, and \( \bar{x} \) is the mean of all split-sample values (33).

On the basis of the excretion kinetics (Figure 2B), which indicated that most urinary excretion of proline betaine occurred in \( \leq 14 \) h, the citrus consumption reported for the morning of the previous day (>20 h before the start of urine collection) was not expected to affect urinary proline betaine concentrations, whereas citrus consumption on the previous evening (<12 h before the start of urine collection) might have resulted in the elevation of urinary proline betaine concentrations (see supplemental Figure 1 under “Supplemental data” in the online issue). Therefore, we categorized reported citrus fruit intake from the 24-h recall data into 3 subgroups as follows: no recorded citrus intake, recorded citrus intake (citrus intake on days 1 and 2 or day 2 only), and recorded citrus intake on day 1 only. Proline betaine concentrations differed significantly between individuals with no recorded citrus consumption and individuals with recorded citrus consumption of proline betaine in the training set and validation set (Belfast sample: \( P < 0.0001 \); West Bromwich sample: \( P < 0.0001 \); Figure 3, A and B). In non-citrus consumers who excreted proline betaine, we observed no consistent pattern of blue or brie cheese consumption or other foods reported to contain low amounts of proline betaine (34). ROC curves for proline betaine were derived for the training and validation sets with an area under the curve (AUC) of 92.3% and 93.5%, respectively (Figure 3C). A cutoff was calculated from the optimal operating point on the ROC curve of the training set with a threshold of 39.4 for the proline betaine integral. This optimal point had a specificity and sensitivity of 90.6% and 86.2% for the training set and 92.3% and 80.6% for the validation set, respectively. To provide further biomarker validation, we repeated these analyses with use of the ratio of proline betaine and creatinine. Sensitivity and specificity were similar for ROC curves.
constructed from the proline betaine:creatinine ratio (89.1% and 83.9%, respectively; AUC: 93.5%); similar results were also obtained for the second rather than the first 24-h urine collection (90.2% and 84.8%, respectively; AUC: 90.7%) (see supplemental Table 1 under “Supplemental data” in the online issue).

Comparison of nutrient intakes, BMI, and blood pressure between citrus fruit consumers and nonconsumers

We classified participants from the INTERMAP US and UK samples as citrus consumers and noncitrus consumers on the basis of their urinary proline betaine excretions. Citrus consumers reported higher intakes of carbohydrates (49.2% compared with 45.5% of energy) and vegetable protein (5.8% compared with 5.4% of energy) and less total fat (31.0% compared with 34.2% of energy) and animal protein (9.7% compared with 10.2% of energy) (all \(P<0.0001\)) compared with noncitrus consumers; intakes of total saturated fatty acids, monounsaturated fatty acids, \textit{trans} fatty acids, polyunsaturated fatty acids, omega-6 fatty acids, and cholesterol were also lower in citrus consumers (Table 2). Most differences were favorable in terms of a healthier diet (ie, a low-fat and high-carbohydrate diet that is rich in fruit, vegetables, and fiber). Citrus consumers ingested more total sugars (25.9% compared with 22.2% of energy) derived mainly from higher fructose (5.2% compared with 3.6% of energy) and glucose (5.3% compared with 3.8% of energy) intakes; fiber and most vitamin intakes (vitamin A, \(\beta\)-carotene, thiamine, pantothenic acid, vitamin B-6, vitamin C, and folic acid) and mineral intakes (copper, iron, magnesium, and urinary potassium) were also higher in citrus consumers, whereas the urinary sodium-potassium ratio was lower (2.9 compared with 2.3) (Table 2). The mean vitamin C intake for consumers was 17.4 mg vitamin C/1000 kJ (150.7 mg vitamin C/d) compared with 7.7 mg vitamin C/1000 kJ (65.5 mg vitamin C/d) for nonconsumers; therefore, as a group, nonconsumers were not meeting the US National Academy of Sciences recommendations for vitamin C intake [75–90 mg vitamin C/d (35)]. In addition, citrus consumers had a higher socioeconomic status assessed by education years (14.5 compared with 13.3 y), lower BMI (27.6 compared with 28.7), and lower systolic blood pressure (118.5 compared with 120.2 mm Hg; \(P<0.005\)) (Table 3).

DISCUSSION

We present a strategy for food biomarker discovery on the basis of untargeted metabolic profiling of urine specimens from a nutritional intervention study and the subsequent validation of

---

**TABLE 1**

Proline betaine concentrations in fruit and fruit juices measured by \(^1\)H nuclear magnetic resonance spectroscopy

<table>
<thead>
<tr>
<th>Juice/fruit description</th>
<th>Proline betaine (mg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citrus fruit juice (authentic; all (n=3))</td>
<td></td>
</tr>
<tr>
<td>Orange juice From concentrate</td>
<td>1316 ± 72^2</td>
</tr>
<tr>
<td>Not from concentrate</td>
<td>1189 ± 24</td>
</tr>
<tr>
<td>Freshly squeezed</td>
<td>1062 ± 81</td>
</tr>
<tr>
<td>Grapefruit juice</td>
<td>766 ± 93</td>
</tr>
<tr>
<td>Citrus fruit juice (synthetic; all (n=1))</td>
<td></td>
</tr>
<tr>
<td>Orange soft drink</td>
<td>216</td>
</tr>
<tr>
<td>Orange squash</td>
<td>75</td>
</tr>
<tr>
<td>Citrus fruit (all (n=3))</td>
<td></td>
</tr>
<tr>
<td>Orange</td>
<td>761 ± 89</td>
</tr>
<tr>
<td>Lime</td>
<td>730 ± 126</td>
</tr>
<tr>
<td>Satsuma</td>
<td>461 ± 55</td>
</tr>
<tr>
<td>Lemon</td>
<td>251 ± 153</td>
</tr>
<tr>
<td>Other fruit juice (all (n=1))</td>
<td></td>
</tr>
<tr>
<td>Pineapple juice</td>
<td>57</td>
</tr>
<tr>
<td>Red grape and raspberry juice</td>
<td>46</td>
</tr>
<tr>
<td>Pomegranate and blueberry juice</td>
<td>18</td>
</tr>
<tr>
<td>Peach, mango, passion fruit juice</td>
<td>17</td>
</tr>
<tr>
<td>Apple juice</td>
<td>14</td>
</tr>
<tr>
<td>Black currant juice</td>
<td>12</td>
</tr>
<tr>
<td>Other fruit (all (n=1))</td>
<td></td>
</tr>
<tr>
<td>Kiwi</td>
<td>66</td>
</tr>
<tr>
<td>Grape</td>
<td>51</td>
</tr>
<tr>
<td>Melon</td>
<td>34</td>
</tr>
<tr>
<td>Banana</td>
<td>28</td>
</tr>
<tr>
<td>Strawberry</td>
<td>22</td>
</tr>
<tr>
<td>Pear</td>
<td>14</td>
</tr>
<tr>
<td>Apricot</td>
<td>10</td>
</tr>
</tbody>
</table>

^2 Mean ± SD (all such values); the SD is indicated when more than one sample was analyzed.
the candidate biomarker by using epidemiologic data. We identify urinary excretion of proline betaine as a specific and sensitive biomarker of citrus fruit intake. Although there are several nutritional biomarkers, such as total urinary nitrogen or urea for protein intake and 24-h urinary sodium and potassium for sodium and potassium intake (30, 36), there are few validated biomarkers for specific foods, the most cited example being increased excretion of resveratrol after wine consumption (37).

Sample sizes used in nutritional intervention studies tend to be small, and dietary interventions are usually given in a controlled setting. Therefore, proposed food or nutrient biomarkers derived from such studies cannot always be extrapolated to population studies in free-living individuals. In the current study, we used the rich data resource of a well-validated large-scale epidemiologic study to establish whether the biomarker for citrus fruit is generalizable to populations and to ascertain whether it can act as a surrogate indicator of healthier eating patterns.

We first identified proline betaine as a candidate citrus fruit biomarker by analyzing urine specimens from a food intervention trial by 1H NMR untargeted metabolite profiling (ie, without preselection and prior knowledge of the metabolites to be measured). Our follow-up kinetics study showed that proline betaine was excreted rapidly in urine, and urinary excretion was nearly complete after 24 h, which indicated that proline betaine is metabolically inert or minimally metabolized in humans, which is in agreement with other reports in the literature (38). Because proline betaine is not metabolized, and the CH3 signal used in the measurement is in a relatively uncrowded spectral region and, therefore, not compromised by other metabolite signals, it is a robust indicator of citrus fruit intake and the quantity of citrus fruit consumed. Thus, the use of proline betaine as a biomarker gives a quantitative and qualitative measure of citrus fruit intake, (see supplemental Figure 2 under “Supplementary data” in the online issue) which is an important advantage over many questionnaire approaches that are qualitative or semiquantitative in nature (39, 40).

Proline betaine is known to act as an osmoprotectant in citrus fruit, alfalfa sprouts (41), molluscs (42), and bacteria (43). Previous reports identified proline betaine in orange juice (44), and in humans, proline betaine concentrations were reported to be increased in plasma and urine after orange juice consumption (45). From analysis of fruit and commercial juices, we ascertained that proline betaine concentrations varied according to the type of citrus fruit and the type of fruit processing (commercially processed compared with freshly squeezed juice), which suggests the importance of, eg, pulp extraction and pasteurization (46). This is consistent with reports that phenolic compounds and vitamin C concentrations increase with more vigorous juice extraction and squeezing techniques (47). We observed a higher (>6 times) proline betaine content in authentic (100%) orange juice than in orange soft drinks and orange squash that allowed us to differentiate between citrus fruit juices and synthetic citrus drinks, which may not always be correctly differentiated by participants in dietary recall studies.

Urinary excretion of proline betaine was also shown to have high specificity and sensitivity as a marker of citrus fruit intake in

### Table 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Predicted noncitrus consumers</th>
<th>Predicted citrus consumers</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy (kJ/24 h)</td>
<td>9214.8</td>
<td>9286.4</td>
<td>0.53</td>
</tr>
<tr>
<td>Total SFA (% of energy)</td>
<td>11.9</td>
<td>10.7</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Total MUFA (% of energy)</td>
<td>12.1</td>
<td>10.9</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Total PUFA (% of energy)</td>
<td>6.9</td>
<td>6.2</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Cholesterol (mg/1000 kJ)</td>
<td>31.5</td>
<td>28.1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Starch (% of energy)</td>
<td>24.0</td>
<td>24.0</td>
<td>0.95</td>
</tr>
<tr>
<td>Total sugars (% of energy)</td>
<td>22.2</td>
<td>25.9</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Fructose (% of energy)</td>
<td>3.6</td>
<td>5.2</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Glucose (% of energy)</td>
<td>3.8</td>
<td>5.3</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Fiber (g/1000 kJ)</td>
<td>2.4</td>
<td>2.7</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Vegetable protein (% of energy)</td>
<td>5.4</td>
<td>5.8</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Animal protein (% of energy)</td>
<td>10.2</td>
<td>9.7</td>
<td>0.0003</td>
</tr>
<tr>
<td>Urinary sodium (mmol/24 h)</td>
<td>155.7</td>
<td>150.7</td>
<td>0.05</td>
</tr>
<tr>
<td>Urinary potassium (mmol/24 h)</td>
<td>59.2</td>
<td>69.2</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Urinary sodium-potassium ratio (mmol:mmol)</td>
<td>2.9</td>
<td>2.3</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Calcium (mg/1000 kJ)</td>
<td>95.5</td>
<td>99.9</td>
<td>0.003</td>
</tr>
<tr>
<td>Magnesium (mg/1000 kJ)</td>
<td>34.9</td>
<td>38.0</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Phosphorus (mg/1000 kJ)</td>
<td>148.8</td>
<td>151.9</td>
<td>0.02</td>
</tr>
<tr>
<td>β-Carotene (μg/1000 kJ)</td>
<td>315.8</td>
<td>437.8</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Vitamin C (mg/1000 kJ)</td>
<td>7.7</td>
<td>17.4</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Folic acid (μg/1000 kJ)</td>
<td>31.2</td>
<td>36.9</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

1 All values are means and were adjusted for age, sex, and country by least-squares means from all 4 visits. SFA, saturated fatty acid; MUFA, monounsaturated fatty acid; PUFA, polyunsaturated fatty acid. Predicted citrus consumers had urinary proline betaine integrals >39.4 in both 24-h collections; predicted noncitrus consumers had urinary proline betaine integrals ≤39.4 in both 24-h collections. Individuals with a urinary proline betaine integral >39.4 in one 24-h collection only (n = 645) were excluded.

---

**FIGURE 3.** Box plots of urinary proline betaine excretion of volunteers recording no citrus fruit consumption (no citrus), citrus fruit consumption (citrus), and citrus fruit consumption only on day 1 (only D 1) A: Proline betaine in the Belfast (UK) sample (no citrus: n = 96; citrus: n = 96; only D 1: n = 28). B: Proline betaine in the West Bromwich (UK) sample (no citrus: n = 181; citrus: n = 71; only D 1: n = 27). C: Receiver operating characteristic curves to assess the predictive ability of excretion of proline betaine for discrimination of citrus fruit intake and no citrus fruit intake as reported in the dietary recall data for the training set (International Collaborative Study of Macronutrients, Micronutrients, and Blood Pressure (INTERMAP) UK Belfast sample) and test set (INTERMAP UK West Bromwich sample). The optimal operating point (●) for the training set was a peak integral value of 39.4 for proline betaine. This represented a specificity and sensitivity of 90.6% and 86.3%, respectively, for the training set and 92.3% and 80.6%, respectively, for the validation set.
high-quality epidemiologic data by using the multipass 24-h dietary recall method as an independent measure of citrus fruit intake (30). Increased excretion of proline betaine after consumption of brie or blue cheese (34) and alfalfa sprouts has also been reported (48) but at substantially lower concentrations (approximately one-fifth the concentration of citrus fruit). Interrogation of the INTERMAP data did not find an association between blue or brie cheese consumption and proline betaine concentrations, and none of the participants in the INTERMAP UK samples reported consuming alfalfa sprouts. Thus in the INTERMAP sample, the measured proline betaine excretion was related overwhelming to citrus fruit intake because other dietary sources did not contribute to the proline betaine concentrations, and proline betaine is not synthesized or metabolized in the lumen.

We observed proline betaine to be a robust biomarker of citrus fruit consumption in Western populations. However, to our knowledge, this biomarker has not been validated in Eastern populations. Furthermore, because proline betaine is predominantly excreted ≤24 h after consumption, it can only be used as evidence of an acute intake of citrus fruit.

Although it is important to ascertain the effects of single nutrients or foods with respect to risk of diseases such as cardiovascular disease or cancers, more recently, there has been a shift in emphasis to address the relatively greater effect of dietary patterns with respect to disease risk. Thus, a food pattern dominated by fruit, vegetables, fish, whole grains, and legumes correlates negatively with attributes of the metabolic syndrome (9), cardiovascular disease risk factors (49), risk of colorectal cancer (10), and blood pressure and serum lipids (12). Our analyses of the INTERMAP UK and US samples showed that there were significant differences in dietary patterns for citrus fruit consumers than for nonconsumers, as confirmed by NMR-detected urinary proline betaine excretion. Citrus fruit consumers had a diet lower in total fats and urinary sodium-potassium ratios, higher intakes of vegetable protein, most micronutrients, and fiber, and lower BMI values and systolic blood pressure measurements, which conferred a lower risk of cardiovascular disease, than did noncitrus consumers (50).

In conclusion, we showed the use of a strategy that combines metabolome-wide association via untargeted metabolic profiling in nutritional intervention studies with validation in free-living populations to identify and validate biomarkers of food intake, which was exemplified in the current study by proline betaine for citrus fruit intake. This nutrimentabonomics approach to biomarker identification and verification should facilitate the evaluation of individual diets toward healthier eating to promote healthy lifestyles and longevity.

We thank TMD Ebbels and H Keun for providing in-house Matlab software and E Maibaum for performing 1H NMR acquisition. We also thank the principal investigators responsible for the INTERMAP UK West Bromwich (G Beevers, G Lip) and Belfast (A Evans, J Yarnell) population sample collections.

The authors’ responsibilities were as follows—SSH: developed the analytical strategy, wrote the manuscript, and conducted the experiment, data analysis, and interpretation of the data; IJB: analyzed the data and wrote the manuscript; QC: developed the analytical strategy and analyzed the data; MB: helped conceptualize the INTERMAP 1H NMR spectral data; M-ED: provided statistical support; SK: helped conceptualize the design of the nutritional studies; JS: designed the INTERMAP study and provided input for writing the manuscript; EH: helped design and implement the NMR data analysis of the INTERMAP samples and provided input for the writing of the manuscript; PE: designed the INTERMAP study, obtained funds for the NMR data analysis of the INTERMAP samples and provided input for the writing of the manuscript; and JKN: obtained funds for the NMR data analysis of the INTERMAP samples and provided input for the writing of the manuscript. None of the authors declared a conflict of interest.

REFERENCES


