A metabolomics approach to the identification of biomarkers of sugar-sweetened beverage intake

Helena Gibbons, Breige A McNulty, Anne P Nugent, Janette Walton, Albert Flynn, Michael J Gibney, and Lorraine Brennan

ABSTRACT
Background: The association between sugar-sweetened beverages (SSBs) and health risks remains controversial. To clarify proposed links, reliable and accurate dietary assessment methods of food intakes are essential. 
Objective: The aim of this present work was to use a metabolomics approach to identify a panel of urinary biomarkers indicative of SSB consumption from a national food consumption survey and subsequently validate this panel in an acute intervention study.
Design: Heat map analysis was performed to identify correlations between 1H nuclear magnetic resonance (NMR) spectral regions and SSB intakes in participants of the National Adult Nutrition Survey (n = 565). Metabolites were identified and receiver operating characteristic (ROC) analysis was performed to assess sensitivity and specificity of biomarkers. The panel of biomarkers was validated in an acute study (n = 10). A fasting first-void urine sample and postprandial samples (2, 4, 6 h) were collected after SSB consumption. After NMR spectroscopic profiling of the urine samples, multivariate data analysis was applied.
Results: A panel of 4 biomarkers—formate, citrulline, taurine, and isocitrate—were identified as markers of SSB intake. This panel of biomarkers had an area under the curve of 0.8 for ROC analysis and a sensitivity and specificity of 0.7 and 0.8, respectively. All 4 biomarkers were identified in the SSB sample. After acute consumption of an SSB drink, all 4 metabolites increased in the urine.

Keywords dietary biomarkers, panel of biomarkers, metabolomics, sugar-sweetened beverages, acute intervention study

INTRODUCTION
Dietary factors, including nutrients and non-nutrients, interact with various ongoing metabolic pathways to reduce or increase the risk of diseases such as cardiovascular disease (CVD), obesity, diabetes, and cancer (1). It is well established that the risk of CVD is closely related to the fat composition of the diet, with SFAs contributing to increased concentrations of LDL cholesterol (2, 3). In contrast, a diet rich in fruit and vegetables, such as the Mediterranean diet, is associated with a reduced risk of CVD (4–7). Therefore, because diet is a key environmental factor influencing disease risk, reliable and accurate dietary assessment methods are essential to strengthen the evidence base for dietary recommendations and to decrease the risk of major chronic disease.

Current self-reporting dietary assessment methods (food-frequency questionnaires, 24-h dietary recalls, weighed food records) are associated with errors that include energy underreporting and recall errors (1, 8). For example, these limitations were shown to affect the accuracy of food-frequency questionnaires, resulting in underreporting of total energy and protein in an obese population (9). Dietary biomarkers offer a more objective measure of dietary intake while also complementing existing dietary assessment methods (10). Such biomarkers exist for salt and protein intakes (sodium and nitrogen measured in 24-h urine samples) and energy expenditure (the doubly labeled water technique) (1, 11). These biomarkers provide nearly unbiased estimates of absolute intake and are therefore extremely useful for validating self-reporting instruments (12).

Metabolomics has emerged as a key tool in dietary biomarker discovery and its application has, in general, taken 3 discovery approaches: 1) acute intervention studies, 2) cohort studies, and 3) analysis of dietary patterns and metabolic profiles. These discovery approaches have led to the successful identification of putative biomarkers of exposure to many foods, including citrus fruit (13–16), salmon (16, 17), red meat (18, 19), cruciferous vegetables (20, 21), and coffee (22, 23). Metabolic profiles that are reflective of...
dietary patterns have also been identified (24–26). The food metabolome and metabolomic dietary biomarkers were recently presented in a review (27).

To date, dietary biomarker discovery has focused on single markers to represent entire food groups. Combinations of biomarkers may provide more accurate measures of dietary exposures. The aim of the present work was to use a metabolomics-based approach to identify a panel of biomarkers associated with sugar-sweetened beverage (SSB) intake. The association between the high consumption of SSBs and disease has been identified in certain studies (28, 29), but findings were not always consistent (30–32). A scientific opinion by the European Food Safety Authority also concluded that additional justification for the correlation between SSB consumption and adverse health effects is required (33). Biomarkers of SSB intake will therefore further increase our understanding of the link between SSB intake and health. A panel of metabolites indicative of SSB intake was first identified by using an untargeted approach and then validated in an acute intervention study.

METHODS

Discovery study

Dietary intake data and 1H nuclear magnetic resonance (NMR) urine spectra from 565 National Adult Nutrition Survey (NANS) participants were used for this analysis. The 565 NANS participants were randomly selected from the main NANS database to ensure that there were equal numbers of men and women across the age range. Ethical approval was obtained from the University College Cork Clinical Research Ethics Committee of the Cork Teaching Hospitals and recruitment began in May 2008. NANS investigated habitual food and beverage consumption, lifestyle, health indicators and attitudes to food and health in a representative sample of 1500 adults aged ≥18 y in the Republic of Ireland during 2008–2010 (34). Dietary intake was determined by using a 4-d semiweighed food record, as previously reported (35). Participants were asked to record detailed information on the amount and type of all foods, drinks, and nutritional supplements consumed over 4 consecutive days in a food diary. Each of the 2552 food codes consumed during the survey were assigned to 1 of 68 food groups. For the purpose of this analysis, this was reduced into 37 food groups (listed in Figure 1). A fasting first-void urine sample was also collected (34).

Acute study: ethical approval, subject recruitment, and study design

The acute study was approved by the Human Research Ethics Committee at University College Dublin (UCD). All procedures were conducted according to the principles expressed in the
Declaration of Helsinki. The study was designed to detect urinary biomarkers of SSB consumption. Ten participants were recruited from UCD. Inclusion criteria included healthy men and women aged between 18 and 64 y, not taking any regular medication (with the exception of contraceptive pills), not having any known chronic or infectious diseases, and not planning a pregnancy or pregnant or lactating. After successful completion of screening, participants provided an informed consent. Participants were instructed not to consume any SSBs 7 d before partaking in the study. SSBs were defined as all nondiet carbonated beverages excluding carbonated alcohols and waters. On the day before sample collection, subjects were instructed to exclude any fish or alcohol from their diet because these can create large spectral peaks that distort multivariate analysis. They were also asked to abstain from medication and to avoid any strenuous physical activity.

On the morning of sample collection, participants collected a fasting first-void urine sample, which was immediately placed on ice and brought to the human intervention suite in the Institute of Food and Health in UCD. Body height was measured by using the Leicester portable height measure (Chasmores), weight was measured by using a Tanita body composition analyzer BC-420MA, and waist and hip circumferences were measured by using a nonstretche tape measure and taken at the naked site where possible. Height and weight were used to calculate BMI (weight [kg]/height [m]^2), and waist and hip circumferences were used to calculate waist-hip ratio [waist [cm]:hip [cm]]. The SSB (330-mL can of cola) was consumed by each participant. Urine samples were collected at 2, 4, and 6 h after consumption. Participants refrained from eating and drinking until the final urine sample was collected. All urine samples were centrifuged at 1800 × g for 10 min at 4°C, and five 1-mL aliquots were stored at ~8°C for NMR analysis (see Supplemental Figure 1 for study design).

Urine data acquisition

Urine samples were prepared by the addition of 250 µL phosphate buffer (0.2 mol K_2HPO_4/L, 0.8 mol K_3HPO_4/L) to 500 µL urine. After centrifugation at 6082 × g for 5 min at 4°C, 10 µL sodium trimethylsilyl [2,2,3,3-^2H_4]propionate (TSP) and 50 µL deuterium oxide (D_2O) were added to 540 µL of the supernatant. Spectra were acquired on a 600-MHz Varian NMR spectrometer by using the first increment of a nuclear Overhauser enhancement spectroscopy pulse sequence at 25°C. Spectra were acquired with 16,384 data points and 128 scans. Water suppression was achieved during the relaxation delay (2.5 s) and the mixing time (100 ms). All ^1H NMR spectra were referenced to TSP at 0.0 parts per million and processed manually with the Chenomx NMR Suite (version 7.5) by using a line broadening of 0.2 Hz, followed by phase and baseline correction. Data were normalized to the total area of the spectra integral. Spectra were exported as 700 and 8500 spectral regions for the discovery study and the acute study, respectively. The water region (4.0–6.0 parts per million) was excluded. Osmolality was measured by using an Advanced Micro Osmometer model 3300 (Advanced Instruments). Aliquots of urine were tested for osmolality with the use of micro-osmometry by freezing point depression, with values reported as the number of solute particles, in moles, dissolved in a kilogram of urine. Metabolite concentrations were normalized to osmolality where appropriate.

To confirm that metabolites were correctly assigned, a solution of pure compound was added to a urine sample. A ^1H NMR spectrum was collected before and after the addition of the pure compounds. Furthermore, a ^1H NMR spectrum of the SSB was acquired. The SSB was degassed by ultrasonication for 10 min, and 500 µL of the degassed solution was combined with 250 µL phosphate. After centrifugation at 6082 × g for 5 min at 4°C, 10 µL TSP and 50 µL D_2O were added to 540 µL of the supernatant. The final solution (600 µL) was transferred into an NMR tube for direct measurement by NMR spectroscopy.

Statistical analyses

Heat map analysis was performed by using regularized canonical correlations analysis (rCCA) in the mixOmics package in R (version 2.15) (36). This multivariate analysis technique allows exploration of correlations between 2 sets of quantitative variables observed on the same experimental units, therefore allowing visualization of correlations between ^1H NMR spectral regions and food group intakes (37). Analysis focused on the SSB food group. Mann-Whitney U test was performed by using IBM SPSS Statistics 20 to compare spectral region intensities and metabolite concentrations between consumers and nonconsumers of SSBs. Multiple comparisons were adjusted by using Bonferroni’s correction. Metabolites for both the discovery study and acute intervention study were identified and quantified by using the Chenomx Profiler (version 7.5).

Multivariate data analyses were performed with Simca-P+ software (version 13.0; Umetrics) within the acute study. Data sets were scaled by using Pareto scaling. Principal components analysis (PCA) was applied to the validation data set to explore any trends or outliers in the data. The differences between NMR spectral data were further explored by using partial least-squares discriminant analysis (PLS-DA). Subsequently, orthogonal PLS-DA (OPLS-DA) was applied, allowing improved interpretation of PLS-DA models because it removes any systematic variation from the data (38). From the OPLS-DA models, an S-line plot was generated. This method gives an indication of the relative influence of metabolites in discriminating between the time points within the acute study.

Receiver operating characteristic (ROC) curves were constructed by using the diegR package in R (version 2.15) (39). ROC curves were produced to determine whether the panel of biomarkers could discriminate between consumers and nonconsumers of SSBs. The classification performance (sensitivity and specificity) of the panel of biomarkers in the identification of consumers and nonconsumers of SSBs was assessed by the AUC. The shortest distance from the optimal point (0,1) to the intersect of the ROC curve was used to measure the optimal cutoff for sensitivity and specificity calculation (40).

RESULTS

Identification of a panel of SSB biomarkers by using an untargeted ^1H NMR approach and rCCA in a national food consumption survey

A subcohort of 565 participants from NANS was included in this analysis; characteristics are described in Table 1. The present study analyzed data from 285 men and 280 women with
a mean (±SD) age of 47 ± 16 yr and a mean (±SD) BMI (in kg/m²) of 27.50 ± 5.11. Heat map analysis performed with the use of rCCA was used to visualize correlations between ^1H NMR spectral regions and food group intakes (g/d) (Figure 1).

The heat map revealed strong positive correlations between SSBs and 49 spectral regions. Participants above the 75th percentile (n = 146) of SSB intake were classified as consumers (83–1155 g/d) and those consuming 0 g/d were classified as nonconsumers (n = 391). For consumers only, the mean daily intake of SSBs was 235.9 g. By using this cutoff approach, only 28 participants (5%) of the total cohort were excluded. The 49 spectral regions that were positively correlated with SSBs were compared between the nonconsumers and consumers (Table 2). The intensity of a number of spectral regions was significantly higher in consumers of SSBs (P < 0.05).

The following metabolites—citrulline, formate, isocitrate, and taurine—were assigned to the spectral regions of interest, and the assignment of each metabolite was confirmed by an adding a solution of pure compound to a urine sample. The metabolites were quantified by using Chenomx Profiler, and concentrations (mmol/L) were compared between nonconsumers and consumers of SSBs. All 4 metabolites were significantly higher in consumers of SSBs (P < 0.05) (Table 3). To eliminate the risk of not capturing all consumers, the metabolites were also compared between nonconsumers and all SSB consumers (including the 5% excluded by using the percentile cutoffs, n = 174). All 4 metabolites remained significantly higher in consumers of SSBs (P < 0.05) (Table 3).

An ROC curve was produced by using partial least-squares regression to determine whether the panel of biomarkers could discriminate between nonconsumers and consumers of SSBs (Figure 2). Within this model, the ROC curve used 5-fold cross-validation. The AUC was 0.8, which represented a good ROC curve according to the accepted classification of biomarker utility (20). The cutoff was calculated from the optimal operating point on the ROC curve, and this had a specificity and sensitivity of 0.8 and 0.7, respectively. The assessment of each biomarker separately revealed an AUC ranging from 0.5 to 0.7.

Validation of the panel of SSB biomarkers in an acute intervention study reveals the presence of the biomarkers

To validate our findings, the panel of biomarkers was investigated in an acute intervention study (n = 10) after the consumption of a standard SSB. Participants were female with a mean (±SD) age of 27 ± 3 y and a normal BMI of 21.03 ± 1.91; characteristics are presented in Table 1. Urine collected at baseline (time point 0) and post–SSB consumption (time points 2, 4, and 6 h) were analyzed by using multivariate statistics. The initial PCA of the ^1H NMR urine data showed 1 outlying sample; this was inspected and revealed high concentrations of trimethylamine N-oxide. For the remaining analysis, this outlying sample was removed and PCA was repeated. The PCA revealed good separation when comparing the post–SSB consumption time points with baseline. An example of the PCA is shown for baseline and the time point 4 h post–SSB consumption (R² = 0.453; see Supplemental Figure 2). PLS-DA was applied to further discriminate between baseline urinary profiles and post–SSB consumption urinary profiles. Of those comparisons, baseline and the 4-h time-point postconsumption model had the highest Q² value (0.738) (Figure 3). Further discriminating information between baseline and the time point 4 h post–SSB consumption was extracted by using an S-line plot generated from an OPLS-DA. The panel of biomarkers was identified on the S-line (see Supplemental Figure 3A, B). All 4 biomarkers were then quantified and normalized to osmolality. Examination of the metabolites over the time points revealed a differential time course. Urinary excretion of formate was gradual and peaked at the time point 4 h postconsumption. Citrulline peaked at the time point 2 h postconsumption; isocitrate peaked at the time point 2 h postconsumption, decreased at the time point 4 h postconsumption, and peaked again at the time point 6 h postconsumption. Taurine also peaked at the time point 6 h postconsumption (see Supplemental Figure 4). Furthermore, the assessment of the NMR spectra of SSB sample revealed the presence of the biomarkers in the sample (see Supplemental Figure 5).

DISCUSSION

In the present study, we used a metabolomics-based approach to identify a panel of biomarkers indicative of SSB intake. This

### TABLE 2

<table>
<thead>
<tr>
<th>Region*</th>
<th>Nonconsumers¹</th>
<th>Consumers¹</th>
<th>P*</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.565</td>
<td>0.360 ± 0.118</td>
<td>0.418 ± 0.148</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3.035</td>
<td>0.491 ± 2.101</td>
<td>10.221 ± 2.098</td>
<td>0.001</td>
</tr>
<tr>
<td>1.985</td>
<td>0.124 ± 0.020</td>
<td>0.134 ± 0.020</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2.055</td>
<td>0.311 ± 0.054</td>
<td>0.336 ± 0.056</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3.235</td>
<td>0.253 ± 0.058</td>
<td>0.282 ± 0.064</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3.825</td>
<td>0.606 ± 0.100</td>
<td>0.637 ± 0.099</td>
<td>0.003</td>
</tr>
<tr>
<td>1.645</td>
<td>0.131 ± 0.040</td>
<td>0.149 ± 0.070</td>
<td>0.024</td>
</tr>
<tr>
<td>1.825</td>
<td>0.108 ± 0.018</td>
<td>0.114 ± 0.017</td>
<td>0.009</td>
</tr>
<tr>
<td>1.145</td>
<td>0.092 ± 0.032</td>
<td>0.111 ± 0.061</td>
<td>0.013</td>
</tr>
<tr>
<td>2.005</td>
<td>0.141 ± 0.020</td>
<td>0.152 ± 0.025</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>1.795</td>
<td>0.105 ± 0.016</td>
<td>0.112 ± 0.020</td>
<td>0.005</td>
</tr>
<tr>
<td>3.225</td>
<td>0.223 ± 0.055</td>
<td>0.245 ± 0.062</td>
<td>0.005</td>
</tr>
<tr>
<td>1.655</td>
<td>0.139 ± 0.040</td>
<td>0.157 ± 0.074</td>
<td>0.052</td>
</tr>
<tr>
<td>1.105</td>
<td>0.060 ± 0.025</td>
<td>0.072 ± 0.039</td>
<td>0.003</td>
</tr>
<tr>
<td>2.035</td>
<td>0.328 ± 0.055</td>
<td>0.349 ± 0.052</td>
<td>0.001</td>
</tr>
</tbody>
</table>

¹All values are means ± SDs. Participants below the 69th percentile (n = 391) of mean daily SSB intake (0 g/d) were classified as nonconsumers and those above the 75th percentile (n = 146) of SSB intake (83–1155 g/d) were classified as consumers. SSB, sugar-sweetened beverage.

*Only a selection of the top ranking regions are presented. These regions represent peaks, and one metabolite can have multiple peaks depending on the chemical structure of the metabolite.

The average normalized intensity (arbitrary units) is presented.

Based on Mann-Whitney U test adjusted for multiple comparisons with the use of Bonferroni’s correction.

### TABLE 1

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Discovery study</th>
<th>Acute study</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>565</td>
<td>10</td>
</tr>
<tr>
<td>Sex, n</td>
<td>280 (F), 285 (M)</td>
<td>10 (F)</td>
</tr>
<tr>
<td>Age, y</td>
<td>47 ± 16¹</td>
<td>27 ± 3</td>
</tr>
<tr>
<td>BMI kg/m²</td>
<td>27.50 ± 5.11</td>
<td>21.03 ± 1.91</td>
</tr>
<tr>
<td>Waist:hip circumference ratio</td>
<td>0.89 ± 0.09</td>
<td>0.75 ± 0.04</td>
</tr>
</tbody>
</table>

¹Mean ± SD (all such values).
The carbon isotope ratio carbon-13 was previously proposed as a biomarker of SSB intake, but correlations identified are modest ($r = 0.345$, $P < 0.01$) (46). A further limitation of using blood carbon-13 as an SSB biomarker is that the carbon in blood occurs primarily in the form of proteins and consequently reflects the carbon-13 of dietary protein sources. To overcome this limitation, the carbon-13 value of alanine, which is synthesized from glucose carbon and therefore more reflective of SBB intake, was also proposed as a marker (47).

Because food groups such as SSBs can comprise a diverse range of compounds, together with the many intrinsic and extrinsic factors that can affect biomarker response, single biomarkers are not always appropriate. This problem is evident when using vitamin C as a single marker of fruit and vegetable intake. Vitamin C amounts can vary greatly between fruit and vegetables; therefore, unless vitamin C is the main bioactive ingredient, these differences will influence inaccurate conclusions and lead to biased results (48). A recent study that identified biomarkers of coffee also found that a combination of markers provided higher sensitivity and specificity compared with using a single marker (22). Therefore, a panel of biomarkers may yield more meaningful results.
results than the use of a single marker. In the present study, the ROC curve analysis of the panel of biomarkers showed that the combination of biomarkers was more effective than any of the biomarkers alone. The panel of biomarkers also had a specificity and sensitivity of 0.8 and 0.7, respectively. This result is acceptable when accounting for the errors associated with self-reporting dietary data, which were used to classify the subjects into consumers and nonconsumers. Sensitivity and specificity > 70% in relation to potential food exposure markers were previously regarded to represent a good level of strength (20). We believe that the identified panel is specific for SSBs as opposed to sugar consumption in general because the total sugar consumption in both groups (consumers and nonconsumers) did not differ significantly. Furthermore, the panel of markers was identified in a sample SSB.

The validation of putative biomarkers in a separate study is essential for their acceptance as biomarkers. This validation step is imperative because factors such as genetic variability, lifestyle/physiologic factors, dietary factors (e.g., nutrient-nutrient interaction), biological sample, and analytic methodology, which may not be present in traditional dietary assessment methods, could skew biomarker measures of dietary intake. In general, the validation of current dietary biomarkers varies widely, with proline betaine being a good example of a well-validated citrus fruit biomarker. The current acute study showed the presence of the biomarkers after the consumption of the SSB in a group of subjects with different demographic characteristics compared with the discovery study. The follow-up kinetics showed that the panel of biomarkers increased after SSB consumption. Two of the markers showed a trend toward returning to baseline values at the 6-h time point. This panel may therefore only be used as evidence of acute SSB intake.

This study has many strengths, including the identification of biomarkers within a national food consumption survey that obtained a rich source of dietary data by using a 4-d semiweighed food diary. An additional strength of this study is the discovery of a panel of biomarkers that performs as a stronger marker of intake compared with a single biomarker. Furthermore, this panel of biomarkers was validated in a separate acute intervention study. The bolus of cola consumed in this acute study was similar in quantity to the mean daily intake of SSBs in consumers within the discovery study. This acute study, however, included 10 female participants, which may be a limitation. Further validation may be needed in other populations. Although cola is the most commonly consumed SSB in Europe and the United States, which was evident in the NANS cohort, further validation and testing with other SSBs may be desired (49).

In summary, this study showed the use of a strategy to identify a panel of biomarkers in a free-living population combined with biomarker validation in an acute intervention study. This panel of biomarkers offers the potential of a more objective measure of SSB intake compared with traditional dietary assessment methods. The use of such biomarkers of exposure in conjunction with traditional dietary assessment methods offers immense potential in future studies. Furthermore, the use of dietary biomarkers can facilitate future research into associations between diet and disease and provide new information about the health effects of food, allowing a better substantiation of health claims while also providing the general public with better dietary advice. Future work will ascertain how to translate these markers for use in nutrition epidemiology.

The authors’ responsibilities were as follows—HG and LB: designed and conducted the research, analyzed data, and wrote the manuscript; and BAM, APN, JW, AF, and MJG: provided essential materials and advice. All of the authors read and approved the final manuscript. None of the authors had a conflict of interest.

REFERENCES


