

# Identification of Urinary Food Intake Biomarkers for Milk, Cheese, and Soy-Based Drink by Untargeted GC-MS and NMR in Healthy Humans

Linda H. Münger,<sup>†</sup> Alessia Trimigno,<sup>‡</sup> Gianfranco Picone,<sup>‡</sup> Carola Freiburghaus,<sup>†</sup> Grégory Pimentel,<sup>†,§</sup> Kathryn J. Burton,<sup>§</sup> François P. Pralong,<sup>§</sup> Nathalie Vionnet,<sup>§</sup> Francesco Capozzi,<sup>‡</sup> René Badertscher,<sup>\*,†,§</sup> and Guy Vergères<sup>†</sup>

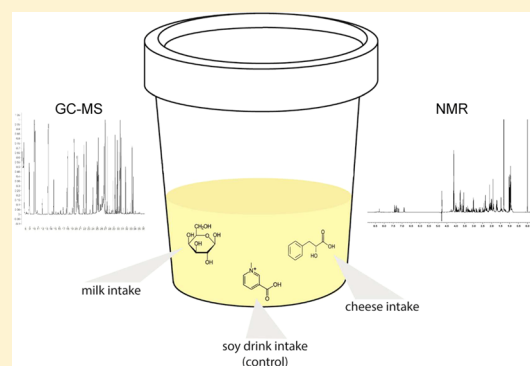
<sup>†</sup>Federal Department of Economic Affairs, Education and Research EAER, Agroscope, Berne, Switzerland

<sup>‡</sup>Department of Agricultural and Food Sciences (DISTAL), University of Bologna, Cesena, Italy

<sup>§</sup>Service of Endocrinology, Diabetes and Metabolism, Lausanne University Hospital, 1011 Lausanne, Switzerland

## Supporting Information

**ABSTRACT:** The measurement of food intake biomarkers (FIBs) in biofluids represents an objective tool for dietary assessment. FIBs of milk and cheese still need more investigation due to the absence of candidate markers. Thus, an acute intervention study has been performed to sensitively and specifically identify candidate FIBs. Eleven healthy male and female volunteers participated in the randomized, controlled crossover study that tested a single intake of milk and cheese as test products, and soy-based drink as a control. Urine samples were collected at baseline and up to 24 h at distinct time intervals (0–1, 1–2, 2–4, 4–6, 6–12, and 12–24 h) and were analyzed using an untargeted multiplatform approach (GC-MS and <sup>1</sup>H NMR). Lactose, galactose, and galactonate were identified exclusively after milk intake while for other metabolites (allantoin, hippurate, galactitol, and galactono-1,5-lactone) a significant increase has been observed. Urinary 3-phenyllactic acid was the only compound specifically reflecting cheese intake although alanine, proline, and pyroglutamic acid were found at significantly higher levels after cheese consumption. In addition, several novel candidate markers for soy drink were identified, such as pinitol and trigonelline. Together, these candidate FIBs of dairy intake could serve as a basis for future validation studies under free-living conditions.



**KEYWORDS:** biomarker, metabolomics, nutrition, urinary metabolome, milk, cheese, soy drink

## ■ INTRODUCTION

Food intake biomarkers (FIBs) are compounds measured in human biological samples, which reflect the intake of specific foods or food constituents.<sup>1</sup> The measurement of FIBs in human biofluids has been proposed as an objective tool to determine intakes of selected food items or food groups, complementing current dietary assessment tools such as food frequency questionnaires or 24 h recalls.<sup>2,3</sup> These traditional assessment tools are susceptible to bias, such as underreporting of dietary intake and inaccurate determination of portion size,<sup>4</sup> that could be overcome by the use of FIBs, which do not rely on self-reporting of food intake. Only a limited number of foods have been associated with FIBs even though a broad coverage of food items is crucial for future applications of dietary assessment. Ultimately, if fully validated, these FIBs could be used for studying the compliance of subjects in nutrimental intervention studies. Moreover, the identification of an accurate and robust set of FIBs, as validation tool for dietary assessments, has been strongly advocated by

epidemiologists to evaluate the nutritional effects of complex diets in case of observational studies.<sup>5</sup>

FIBs can be either single compounds or combined biomarkers, for instance proline betaine is considered a biomarker of citrus fruit consumption<sup>6,7</sup> while tartrate together with ethylglucuronide have been proposed as paired biomarkers of red wine consumption.<sup>8</sup> The ideal FIB is highly specific to one food item or food group, is not detected in the sample of interest when the specific food item is not ingested and shows a distinct dose- and time-dependent response after its intake.<sup>9</sup> Moreover, it should not show susceptibility to interindividual variations. For this reason, most optimal FIBs could be nonmetabolized compounds, as they will be found in urine also several hours after food intake and become suitable for validation of dietary questionnaires.

In the past decade, an increasing number of short- and medium-term dietary intervention studies have been carried out

Received: May 19, 2017

Published: July 28, 2017

in order to identify FIBs that reflect the intake of selected food items, days or weeks after their ingestion.<sup>10,11</sup> These identified FIBs then have to be further validated in cohort studies. The assessment of the sensitivity and specificity of a compound for discriminating between consumers and nonconsumers of the associated food is a prerequisite to ultimately label it as validated FIB.<sup>12</sup>

The consumption of milk and cheese has not yet been characterized by fully validated FIBs even though they are of public health relevance due to the substantial quantities consumed in the Western diet. The main reason for this gap in the field is the lack of specific candidate FIBs for these foods. Intervention and observational studies have identified two plasma fatty acids (C15 and C17), among other lipids, as indicative of the intake of dairy products, in particular dairy fat.<sup>13–15</sup> However, studies in which milk and cheese are investigated as separate entities are rare and the above findings remain controversial due to the issue of specificity.<sup>16</sup> Previous medium-term intervention studies suggest that several non-specific compounds are reflective of milk intake such as serum short chain fatty acids,<sup>17</sup> serum amino acids such as glutamine, proline and aspartic acid<sup>18</sup> and urinary citrate.<sup>19</sup> Similarly, Hjerpsted et al. showed that the presence of amino acid metabolites in urine is indicative of cheese intake compared to butter intake,<sup>20</sup> such metabolites could be the products of either fermentation or human metabolism. Zheng et al. proposed urinary proline betaine and tyrosine as candidate markers for cheese intake,<sup>19</sup> a finding that has to be critically evaluated in light of the proposed identification of proline betaine as a marker of citrus fruit consumption.<sup>7</sup>

Acute intervention studies, which have the ability to sensitively and specifically identify candidate biomarkers, have not yet been performed with milk and cheese. Thus, in the frame of the HDHL Joint Programming Initiative European project “Food Biomarker Alliance” (FoodBALL) (foodmetabome.org), a randomized, controlled, crossover study has been carried out in order to identify novel candidate FIBs for milk and cheese intake compared to a soy-based drink in healthy subjects. Even though soy drink was technically the control food for milk and cheese, candidate markers for this product, which is regularly consumed worldwide, also deserved investigation within this study. Urine samples were collected postprandially at intervals up to 24 h and were analyzed using a multiplatform approach that applied untargeted GC-MS and NMR in order to allow a thorough identification and broad coverage of different groups of metabolites.

## EXPERIMENTAL SECTION

### Study Population

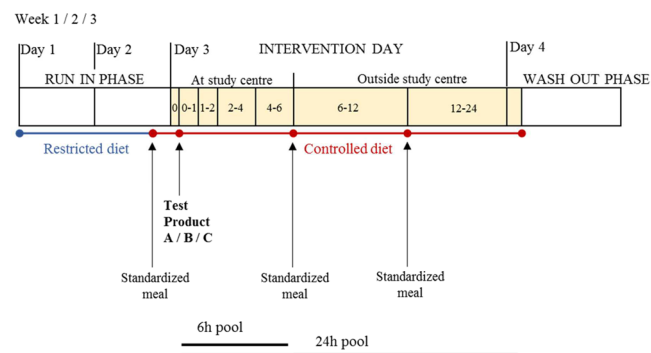
Eleven volunteers were included in the study. All participants signed an informed consent form. The study was conducted according to the guidelines laid down in the Declaration of Helsinki, was approved by the Ethical Committee of the Canton of Vaud (Switzerland) and was registered at clinicaltrials.gov (NCT02705560).

Recruitment was carried out by poster campaign. Volunteers were screened by a telephone interview and invited for a first visit, during which health status, anthropometrics and dietary habits were assessed. In order to exclude adverse effects due to lactose intolerance during the study, volunteers completed a pretest prior to enrolling in the study that required the consumption of 600 mL milk after an overnight fast and signs

of lactose intolerance were assessed over the subsequent 2 h. Inclusion criteria were good health, aged from 18 to 40 years, BMI from 18 to 30 kg/m<sup>2</sup>, regular intake of dairy products and no adverse effects after consumption of 600 mL milk during the pretest. Participants were excluded according to the following exclusion criteria: smoking, diagnosed health condition (chronic or infectious disease), intake of medication (except birth control pills that were permitted), intake of nutritional supplements, pregnancy, lactation, antibiotics treatment within 6 months prior to study, food allergies or intolerances, special diets (vegetarians, nonconsumers of soy), hemoglobin <120 µg/L, ferritin <30 µg/L, and unwillingness to follow dietary restrictions or the controlled diet during the study days. Volunteers were also excluded if they donated blood or participated to another clinical study within 2 and 6 months prior to study, respectively.

### Study Design

The randomized, controlled, crossover study is part of the FoodBALL project running a series of seven acute intervention studies with a broad range of common foods. The design of this study was therefore defined within the FoodBALL project and individually adapted for each study center (Figure 1). During



**Figure 1.** Design of randomized, controlled, crossover study; test products were administered in random order (A = milk, B = cheese, C = soy drink), urine samples were collected before and after ingestion of test product during defined intervals (yellow blocks) up to 24 h; 6 h and 24 h pools were later prepared in the laboratory.

the two-day run-in phase, volunteers followed a restricted diet that required omitting all dairy products, products containing dairy ingredients, fermented products, products from bovine sources and soy products from their diet. Volunteers received dietary advice to support application of the dietary restrictions and kept a food diary to support adherence with the advice. In the evening prior to the day of the intervention, the volunteers consumed a standardized meal at the study center (*ad libitum*) and remained in the fasted state thereafter (minimum 12 h fasting prior to the intervention). The following day the volunteers were advised to urinate at home and drink two glasses of water before coming to the center. Baseline urine was collected at their arrival at the research center (0 h). Each volunteer was assigned to the three different test products in random order. After the consumption of the test product (duration of consumption did not exceed 15 min), urine samples were collected during the time intervals 0–1, 1–2, 2–4, and 4–6 h at the study center. Urine from the time intervals 6–12 and 12–24 h was collected by the participants outside the study center. Participants were not allowed to consume any food or beverages except water (at max. 250 mL per h) during

the 6 h after intake of the test product, during which they remained seated and were only permitted to do light work or walk to the toilet. The volunteers consumed the same standardized meal for lunch and dinner of the intervention day. The lunch was consumed at the study center 6 h after receiving the test product while the dinner was consumed under free-living conditions.

The collected urine samples were stored at 4 °C until further processed. The total volume of each interval collection was recorded, samples were transferred into smaller tubes, centrifuged at 1800g for 10 min at 4 °C, aliquoted, stored overnight at –20 °C and transferred to –80 °C for long-time storage.

### Test Products and Standardized Meal

All test foods were isocaloric (1674 kJ/400 kcal) and isovolumetric (600 mL); 100 g cheese was served along with 500 mL of still water. Pasteurised milk (3.9% fat, organic, brand: Naturaplan) was bought at a local grocery store. One round of hard cheese (Le Gruyère AOP, 7 months ripening time) was obtained from a local cheese maker. The soy drink was composed of a pure soy drink (540 mL, ingredients: water, soya beans, brand: Soy Line) and soy-based vegetable cream (60 mL, ingredients: water, hydrogenated palm oil, glucose, soy flour, emulsifiers (E 475, soya lecithin, E 472b and E 435), thickeners (E 407, locust bean seeds and E 461), stabilizer (E 339) and salt) from the same brand. Test product compositions per serving are presented in Table S-1. The standardized meal was composed of chicken breast, white long grain rice, margarine and salt. The meals were precooked in the study kitchen, packaged into 1200 kcal portions and were served *ad libitum*.

### GC-MS

**Preparation of Pools, Interval Samples, Test Product, and QC Samples.** GC-MS analyses were performed on urine samples collected at baseline and during selected time intervals (referred to as “interval samples”). In addition, two pooled samples were prepared by mixing urine samples after the ingestion of one food by one subject for up to 6 h (6 h pool) and 24 h (24 h pool). For pooling, volumes were combined proportionally to individual volumes of each interval collection. Baseline samples were not included in 6 h and 24 h pools.

Interval samples were prepared with equal specific gravity (1.0008) due to large differences in concentrations. The prenormalization procedure has been discussed elsewhere<sup>21</sup> and was applied with minor modifications. The specific gravity of each urine sample was determined based on refractive index analyses (refractometer RE40, Mettler Toledo, Switzerland).

For cheese analysis, 2.4 g of solid food was mixed with 20 mL of Milli-Q water and the supernatant was collected after centrifugation. All test product samples (100 µL) were precipitated with cold methanol (300 µL) prior to derivatization.

Quality control (QC) samples were prepared from interval samples, 6 h pool, and 24 h pool urine samples as well as cheese, milk and soy drink samples, by mixing all samples of each sample type at equal volume.

**Sample Analysis.** Interval urine samples from each volunteer ( $n = 21$ ) were measured in a randomized order within a single batch. The order of the batches ( $n = 11$  corresponding to the 11 volunteers) was randomized too. In case of pool samples, all 6 h pools ( $n = 33$ ) were randomized and analyzed within one batch followed by the analysis of

randomly ordered 24 h pools ( $n = 33$ ). The test product samples were randomized within one group. Each batch was initiated with 2–5 QC of specific type of samples for equilibration and blank samples at beginning and end of sequence were included. After every fifth sample, a fresh QC sample was injected.

The derivatization was performed in a batch-wise manner. To 100 µL of each sample, 50 µL of internal standard (ISTd) solution (0.17 mg/mL) containing isotopically labeled D-fructose (U-13C6, 99%, Cambridge Isotope Laboratories, Inc., UK) were added. Samples were further dried in SpeedVac at 25 °C for 2 h. The two-stage derivatization and further steps were performed according to the HUSERMET protocol.<sup>22</sup> Blanks (100 µL Milli-Q water) were treated in the same way as other samples.

Urine samples were analyzed on a GC-MS 7890B/MS5977A (Agilent Technologies, Santa Clara, U.S.) and a CombiPAL autosampler (CTC-Analytics AG, Zwingen, Switzerland). The installed liner contained deactivated wool (4 × 6.5 × 78.5 mm, Restek Corporation, Bellefonte, U.S.). One µL of the sample was injected in splitless mode until 1 min then split 1:20. Helium was used as carrier gas at a flow rate of 0.9 mL/min. The separation was performed on a DB-5ms fused silica capillary column (60 m, 0.25 mm i.d., 0.25 µm film thickness) supplied by Agilent Technologies, Basel, Switzerland. The oven program was as follows: initial temperature 70 °C for 2 min, increase up to 160 °C at a rate of 5 °C/min, increase to 300 °C at a rate of 10 °C/min, which was held for 16 min, equilibration time 1 min. MS detection mass ranged from 28.5 to 600 Da, MS source temperature was 230 °C and MS Quad temperature was 150 °C. Electron ionization was performed with 70 eV. The total run time was 50 min for interval samples and 70 min for pools (longer hold time of 36 min). For pools, an injector temperature gradient was applied: initially 90 °C, heating rate 900 °C/min until 280 °C, cooled for 5 min at rate of 30 °C/min and kept at 250 °C. GC-MS analyses of foods were performed on a GCMS 7890A/MS 5975C (Agilent Technologies, Santa Clara, U.S.). GC-MS settings for food samples were as those described above for the pooled urine samples.

**Data Processing, Candidate Marker Selection by OPLS-DA, and Identification of Compounds.** Agilent data files acquired from GC-MS analysis were converted into mzXmL files using ProteoWizard.<sup>23</sup> Data files were further processed in the R environment (version 3.2.2) using the package eRah (version 1.0.4)<sup>24</sup> performing preprocessing, deconvolution, alignment, missing compound recovery and compound identification. The following parameters were used in this data processing:  $\text{min.peak.width} = 2.5$ ,  $\text{avoid.processing.mz} = \text{c}(28.5:69,73:75,147:149)$ ,  $\text{min.spectra.cor} = 0.90$ ,  $\text{max.time.dist} = 2(\text{pools})/2.5(\text{interval samples})$ ,  $\text{mz.range} = 70:600$ . A minimum criteria of 5 samples was applied for the missing compound recovery step. For the first identification of the compounds, the Golm Metabolome Database (GMD)<sup>25</sup> was used.

Multivariate analysis using SIMCA-P software (V.14.0, Umetrics, Umea, Sweden) was performed on the deconvoluted data. The data was mean-centered and UV scaled. Principal component analysis (PCA) was used to detect strong outliers, which are defined as observations clearly falling outside Hotelling's  $T^2$  tolerance ellipse (95% confidence interval) in the score plot. Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA) was used on the 6 h and 24 h pools to differentiate two groups: milk vs cheese/soy drink,



cheese vs milk/soy drink and soy drink vs milk/cheese. Discriminating compounds were selected on loading plot (separated features from rest of features) being the ones with highest Variable Importance in Projection (VIP) values. For the selection of cheese markers, an additional OPLS-DA analysis was conducted on interval samples 0–1, 1–2, 2–4, and 4–6 h and using IStd D-Fructose Peak 1 normalized responses. Only models with predictive ability parameter  $Q^2 > 0.3$  were considered. As this limit falls below the recommended value of 0.5, permutation tests ( $n = 999$ ) were performed to rule out random separation, which is recommended in such cases.<sup>26</sup> In case of stereoisomeric forms of selected discriminating features the peak with higher response was further evaluated. Discriminating features with identical retention time (RT) and showing significant correlations ( $r^2 > 0.7$ , Pearson correlation) were considered as one compound whose spectra was split into multiple IDs during deconvolution.

Each candidate marker was attributed to the level of its identification.<sup>27</sup> Level 1: compounds were identified by comparison to a pure reference (based on spectral data and retention indices (RI)) (Figures S-3 to S-15). Level 2: identification was based on spectral database (match factor >80%). Level 3: only compound groups were known, e.g. specific ions and RT regions of disaccharides. Level 4: unknown compound.

**Targeted Evaluation of Candidate Markers.** All markers selected based on deconvoluted data were further evaluated using a targeted approach in order to optimize integration. Using RI, quantifier and qualifier ion retrieved from deconvoluted data (Table S-3), the suggested markers from the OPLS-DA models were analyzed in MassHunter Quantitative Analysis (Agilent Technologies, Santa Clara, U.S.) in 6 h pools, 24 h pools and QC samples. The peak integration was checked in each sample individually. Responses from the quantifier ion of marker compounds were normalized with the response of the quantifier ion of IStd D-Fructose Peak 1 (ion 279). Univariate analysis using Kruskal–Wallis test was performed in R to confirm significant difference of suggested marker between the foods.  $p$ -values < 0.05 were considered as significant. For multiple comparison, Conover–Inman test (library Conover in R) was performed. QC samples were used to determine coefficient of variance (CV) of each significant candidate biomarker (Table S-3). Significant candidate markers were then categorized into (1) specific markers (above detection limit after the intake of only one of the three foods in 6 h or 24 h pool) or (2) nonspecific markers (above detection limit after intake of more than one out of the three foods but a significant difference was observed between the foods in 6 h or 24 h pool). If a marker was significant in both pools, the pool in which it was more significant was used for further analysis. Statistically significant markers in pools were then further evaluated in interval samples using the targeted method described above and responses normalized to the IStd and corrected for dilution of the samples were calculated as follows: corrected normalized response [a.u.] = response from analysis (normalized with IStd)  $\times$  dilution factor  $\times$  volume of interval. A hypothetical excretion rate was calculated for each significant markers in each interval sample by dividing the corrected normalized response by the duration of the collection interval [h].

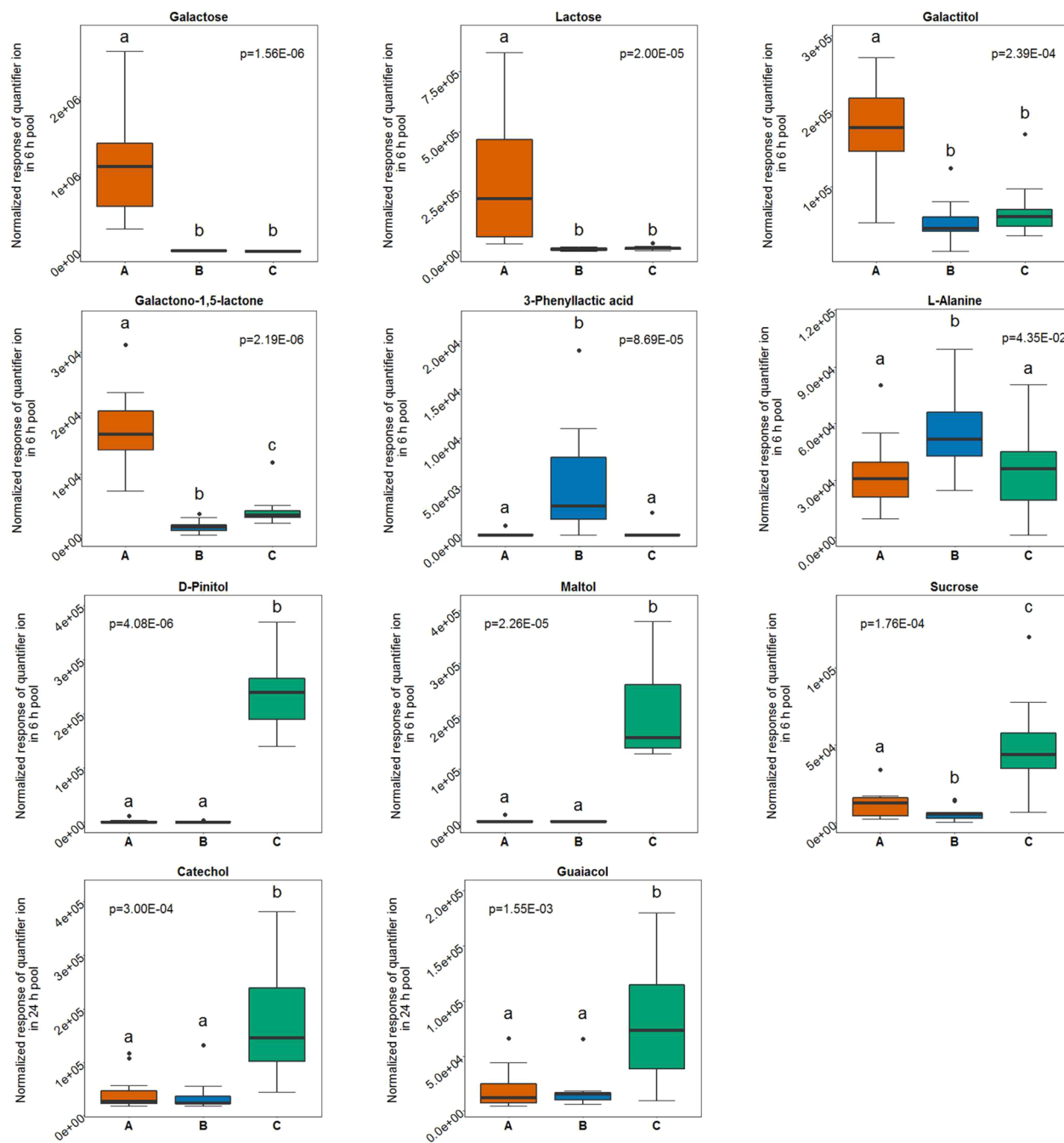
## NMR

**Sample Preparation and Analysis.** Urine aliquots were stored at  $-80\text{ }^\circ\text{C}$  until their use for the NMR analysis. 630  $\mu\text{L}$  of urine sample were centrifuged to remove debris, then 540  $\mu\text{L}$  of supernatant were placed in a clean Eppendorf containing 60  $\mu\text{L}$  of  $\text{D}_2\text{O}$ -based phosphate buffer containing also trimethylsilyl propionate (TSP) as IStd and sodium azide ( $\text{NaN}_3$ ) as an antibacterial agent. A total of 590  $\mu\text{L}$  of the mixture was transferred into  $5 \times 178\text{ mm}$  (7") 5 mm, outer diameter NMR tubes (for Bruker Match holder).  $^1\text{H}$  NMR spectra were recorded at 298 K with an AVANCE spectrometer (Bruker BioSpin, Karlsruhe, Germany) operating at a proton frequency of 600.13 MHz, equipped with an autosampler with 60 holders. The HOD residual signal was suppressed by applying the NOESYGPPRID sequence (a standard pulse sequence included in the Bruker library) incorporating the first increment of the NOESY pulse sequence and a spoil gradient. Each spectrum was acquired using 32 K data points over a 7211.54 Hz spectral width (12 ppm) and summing up 128 transients. A  $90^\circ$  pulse of 12.5  $\mu\text{s}$  was set up. A delay of 5 s between transients, extending the acquisition time of 2.27 s, was chosen to provide a recycle time 5 times longer than the longitudinal relaxation time of the protons under investigation, expected to be not longer than 1.4 s. The data were Fourier transformed and phase and baseline corrections were automatically applied using TopSpin version 3.0 (Bruker BioSpin, Karlsruhe, Germany). Signals were assigned by comparing their chemical shift and multiplicity with Chenomx software data bank (version 8.1, Edmonton, Canada).

**Analysis of Spectra.** Spectra were exported in ASCII file format and then imported into R software (version 3.3.2). Chemical shift referencing was performed by imposing the TSP signal to 0.00 ppm. The spectral regions including only noise (e.g., the spectrum edges below 0.5 and above 10 ppm), as well as the data points which are strongly affected by the residual water (between 4.95 and 4.7 ppm) and the urea signals (5.45–6.1 ppm) were removed prior to data analysis. Normalization was carried out using the PQN algorithm.<sup>28</sup>

For each subject, spectra from samples at the following time points were summed, integrating each point of the selected spectra for each individual, in order to mimic a 6 h urine pool (0–1, 1–2, 2–4, and 4–6 h) and a 24 h pool (0–1, 1–2, 2–4, 4–6, 6–12, and 12–24 h). Only samples with 4 and 6 time points respectively were kept. These spectral matrices were then imported in MatLab (R2014b, MathWorks). In total 95 signals, significantly above the baseline ( $\text{S/N} > 3$ ), were selected, then their integral areas were measured.

These integrals were then imported into R, where OPLS-DA was carried out through the “biocLite” package (<https://www.bioconductor.org/>) using mean centering and unit variance scaling. Treatments were compared 1 by 2, i.e., milk vs soy drink/cheese, cheese vs soy drink/milk, soy drink vs milk/cheese; also comparisons with a 1 by 1 pairwise design were carried out to test other possible classifying models with the same set of foods. A thousand permutations were employed for all models. Signals characterizing most of the variance of the model were then selected: the loadings with greater values than the standard deviation were identified. These signals, most significant for the discrimination, were then analyzed through univariate statistical analysis by means of Kruskal–Wallis with pairwise comparison (through the command “kruskal” in the R package “agricolae”).



**Figure 2.** Boxplots of identified compounds assessed by GC-MS that are significantly different in 6 h pools or in 24 h pools after milk (A), cheese (B), and soy drink (C) intake. The y-axis values are IStd normalized responses of the quantifier ion of each FIB in the respective pool. Compounds are classified into specific (galactose, lactose, 3-phenylactic acid, D-pinitol, maltol) and nonspecific (galactitol, galactono-1,5-lactone, L-alanine, sucrose, catechol, guaiacol); *p*-value based on Kruskal–Wallis test, different letters indicate significant differences between foods based on Conover–Inman test.

For the QC of samples, 3 spectra (one for each sample group) were acquired at three different times. Each of those spectra was then preprocessed by three different operators three times each. The integral of selected biomarkers for the three food were measured on each spectrum. Thus, 27 different measures for the metabolites were made for each of the three selected spectra.

#### Intersubject Variability

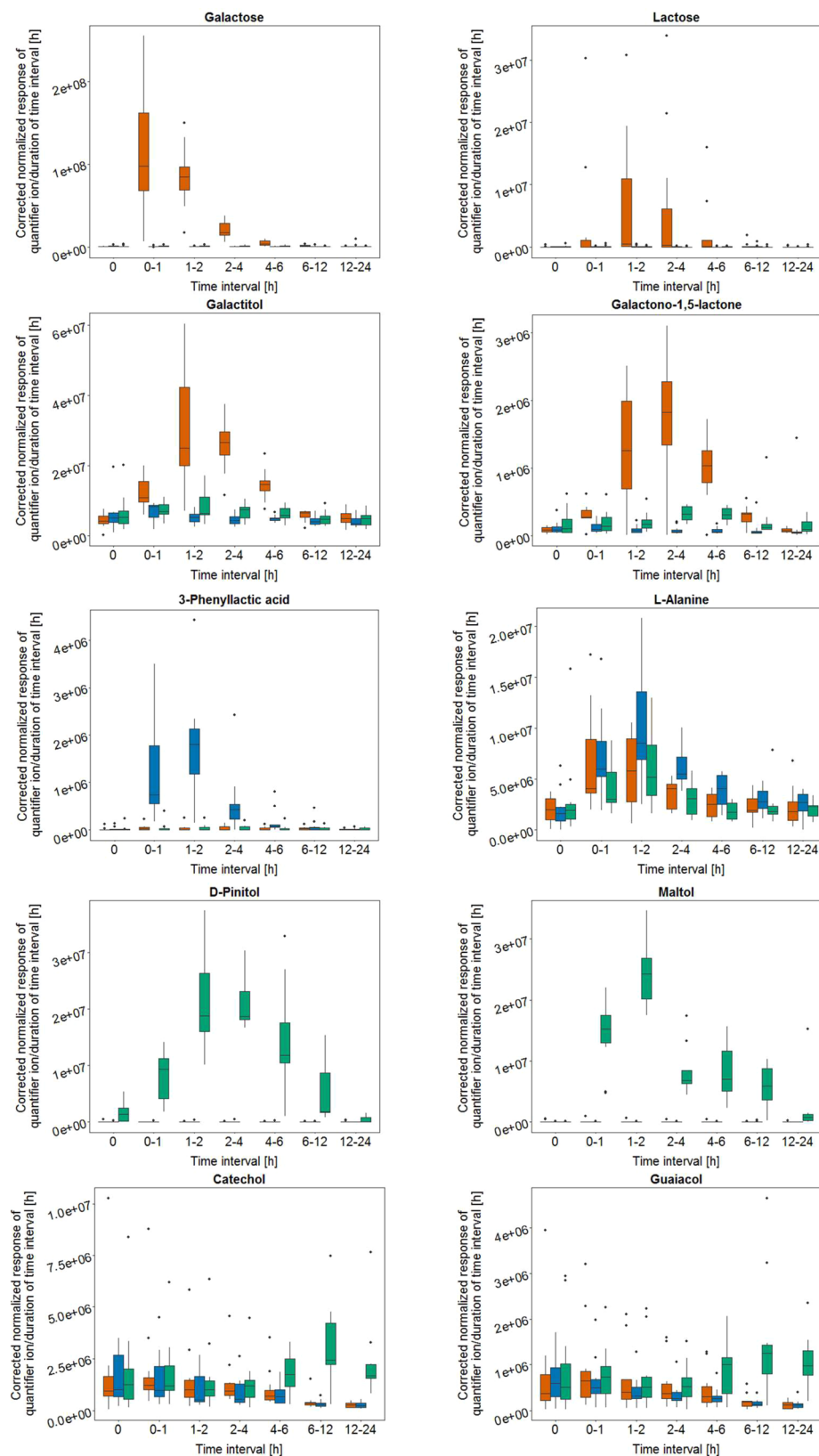
For both analytical methods, percentage CV of metabolites in QC was calculated, together with intersubject percentage CV in

pools and their ratio was calculated. As suggested by Dunn et al., if this factor is greater than >1.5 the metabolic feature is likely to contain biological information.<sup>29</sup>

## RESULTS

### Study Population Characteristics and Test Product Composition

All 11 volunteers, five women and six men, completed the study; details of the study population characteristics are presented in Table S-2. There was one missing urine sample



**Figure 3.** Kinetics of selected discriminating compounds after milk (A), cheese (B), and soy drink (C) intake identified by GC-MS. Y-axis values represent a hypothetical extraction rate of volume adjusted and IStd normalized responses of quantifier ions per duration of time interval for each FIB.

(after milk intake, 6–12 h). The three test products consumed were isocaloric while they differed in the macronutrient composition (Table S-1) with cheese being highest in fat (33

g/100 g) and protein (26 g/100 g) among the three products. The full-fat milk contained 27.5 g of lactose at the volume served (600 mL). Even though each volunteer received a new

batch of pasteurized milk, the macronutrient amounts, dry matter and ash of the milks ( $n = 11$ ) only differed minimally ( $CV < 3\%$ ). Among the sugars analyzed in soy drink, sucrose was highest (5.6 g in 600 mL).

### Candidate FIBs Identified by GC-MS

**Marker Selection Based on OPLS-DA.** In total, respectively 2119, 507, and 662 features were detected after analysis by GC-MS and deconvolution of urinary interval samples, 6 h pools and 24 h pools. During the process of derivatization, artifacts and multiple derivatization products of one compound (multiple TMS derivatives or stereoisomers) are formed, thus the actual number of compounds is lower than the number of features indicated here. In the model based on the interval samples, three outliers were removed based on the PCA score plot. No outliers were identified for pool models based on PCA. There was no need to filter the data as goodness-of-fit parameter ( $R^2$ ) of all OPLS-DA models were higher than 0.95 indicating that the models were not dominated by a large amount of noise.

A summary of all OPLS-DA models (score plot, loading plot, permutation plot and model parameters) used for marker selection is presented in the [Supporting Information](#) (Figures S-1 to S-2, Table S-4). In case of milk and soy drink, the 6 h or 24 h pool models were effective in identifying FIBs, the interval samples were not investigated for FIBs but only for kinetics. For milk and soy drink intake, 6 h pools models resulted in better predictability (milk intake:  $Q^2 = 0.566$ , soy drink intake:  $Q^2 = 0.556$ ) than 24 h pools models (milk intake:  $Q^2 = 0.43$ , soy drink intake:  $Q^2 = 0.325$ ). For cheese intake, the models using 6 h pools ( $Q^2 = 0.426$ ) and interval samples ( $Q^2 = 0.729$ ), but not the 24 h pools, had  $Q^2 > 0.3$  with the interval sample model having highest predictability. In all cases,  $R^2$  and  $Q^2$  exceeded the values generated by the corresponding permutation tests.

Based on the model comparing 6 h pools after milk intake to 6 h pools after cheese and soy drink intake as one group, five compounds, namely galactose, galactitol, lactose and galactono-1,5-lactone (all level 1) and one unknown compound (M10), were identified as the most discriminating metabolites and therefore indicative of milk intake in the 6 h pool. Applying the same analytical approach to the 24 h pool samples, two compounds (galactose and galactitol) were indicative of milk intake.

Five features that showed the greatest discrimination for cheese intake were selected based on the 6 h pool model. Of these features, three eluted very early in the chromatogram, a region in which the most volatile compounds appear but also where excess reagents elute, and thus could not be further evaluated. Therefore, only two features remained - alanine (level 1) and lactic acid (level 2). The additional model using interval samples resulted in the selection of alanine (level 1) in addition to 3-phenyllactic acid (level 1), pyroglutamic acid (level 1) and 4-methylcatechol (level 1).

Models for the selection of markers reflective of soy drink intake based on 6 h and 24 h pools yielded candidate markers, some only being discriminative in one of the pools. D-Pinitol and maltol were identified at level 1 and were markers in both pools similarly to a sugar alcohol (S11, level 3), of unknown identity. Sucrose (level 1), an inositol-like compound (level 2) and three unknown disaccharide features (S12, S13, and S14) were only suggested to be candidate markers based on 6 h pool model. On the other hand, catechol (level 1), guaiacol (level 1)

and an unknown compound (S17) were exclusively indicative of soy drink intake in the 24 h pool analysis.

**Targeted Evaluation of the Candidate FIBs.** The response of the compounds preselected on the basis of OPLS-DA were further evaluated by a targeted approach in the pools and subjected to univariate analysis (Table S-4). The responses of significant specific and nonspecific candidate markers, identified at level 1, for milk, cheese or soy intake in either 6 h or 24 h pools are depicted in Figure 2 (level 3 and 4 compounds are shown in Figure S-16).

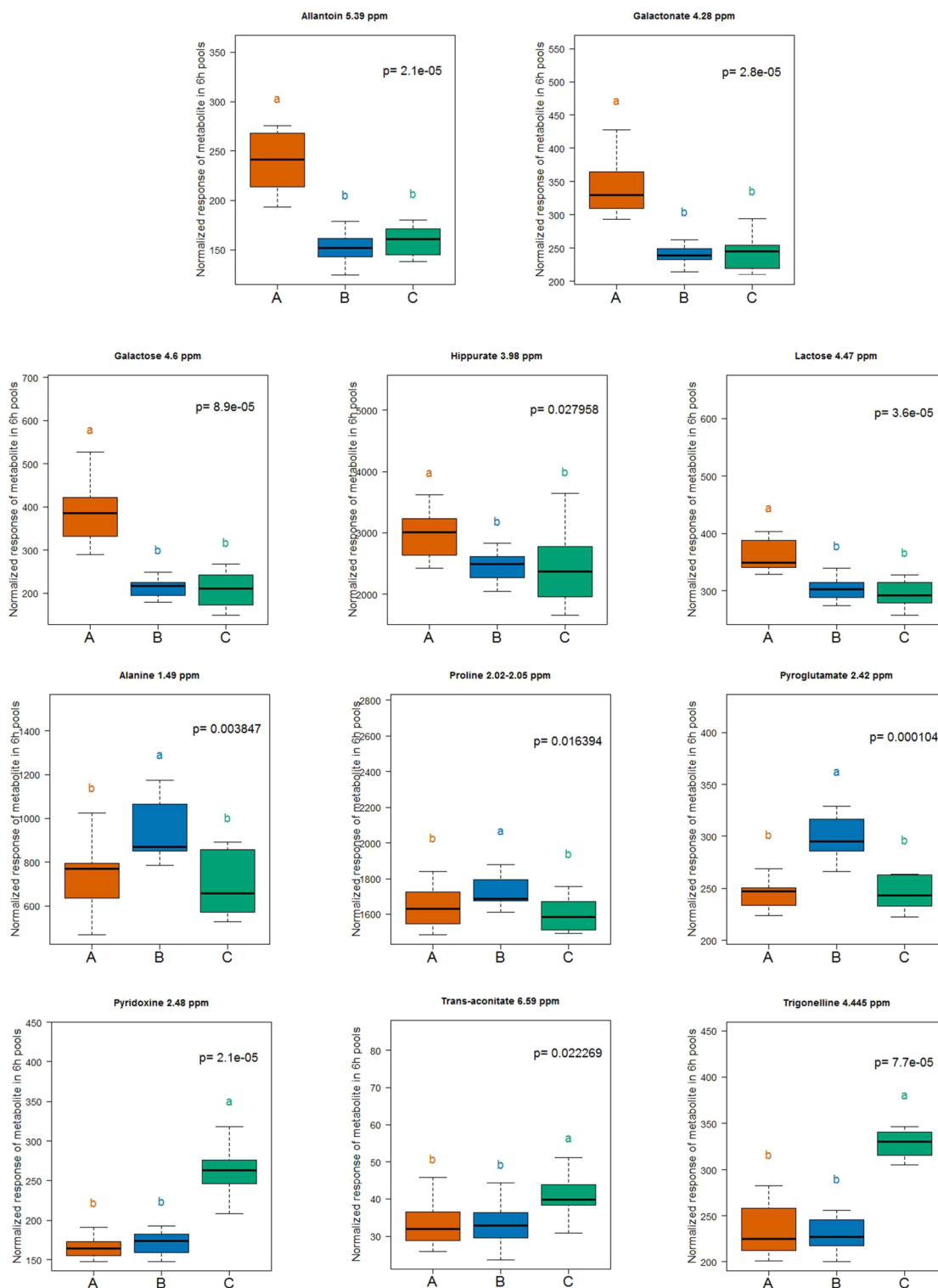
Out of the compounds preselected by OPLS-DA that are reflective of milk intake, all showed significant univariate differences between the three groups. Among these markers, only galactose and lactose were specific for milk intake given that they were below detection limit after the ingestion of the two other products. Galactitol and galactono-1,5-lactone were detected in pooled samples taken after cheese and soy drink intake, however, the levels after milk intake were significantly higher. In the 6 h pools, cheese intake was characterized by significantly higher levels of alanine and 3-phenyllactic acid, the latter being specific to cheese intake. Lactic acid, pyroglutamic acid and 4-methylcatechol, which were selected as candidates with the OPLS-DA approach, were not significant with this targeted evaluation. All compounds preselected by OPLS-DA for soy drink intake were significantly higher by univariate analysis after soy drink intake. However, only D-pinitol and maltol were specific to soy drink among the three foods. All other marker candidates also appeared after the intake of the other two foods but at significantly lower levels.

**Kinetic Analysis of the Candidate FIBs.** The kinetics of selected significant candidate FIBs are shown in Figure 3. Among candidate FIBs for milk, galactose showed the highest excretion rate during 0–1 h while lactose and galactitol reached highest excretion rate 1–2 h and galactono-1,5-lactone 2–4 h after ingestion. By 6 h, these compounds were only present at very low levels in urine. 3-Phenyllactic acid, the specific candidate FIB for cheese intake, reached a maximum excretion rate at 1–2 h while by 4–6 h the excretion rate was already minimal. On the other hand, a clear postprandial response was observed for alanine after the intake all three foods but the excretion rate tended to be higher after cheese intake during 1–24 h. For soy drink the candidate FIBs, D-pinitol and maltol were excreted at maximum rate during 1–2 h and 2–4 h, respectively, but their levels were also elevated in urine samples collected during 6–12 h. A different excretion pattern was found for catechol and guaiacol after soy drink intake; their levels increased only after 4 h.

### Candidate FIBs Identified by NMR

**Spectral Analysis and Targeted Evaluation of Candidate Markers.** The 6 h pool spectral data set consisted of 33 spectra: 11 for milk, 11 for soy drink and 11 for the cheese intervention. In total 95 bins containing signals with  $S/N > 3$  were selected: 34 metabolites, some with multiple signals, were identified, 41 signals remained unassigned (Table S-6). OPLS-DA comparing treatments 1 by 2 on the set of 95 bins resulted in  $Q^2$  values of, respectively, 0.812 (milk vs cheese/soy drink), 0.762 (cheese vs milk/soy drink) and 0.766 (soy drink vs milk/cheese), proving a very good predictability (Figure S-17). A total of 56 variables (bins) had OPLS-DA loadings with values greater than the standard deviation. These signals, most promising for the discrimination, were then analyzed through univariate statistical analysis by means of Kruskal–Wallis test





**Figure 4.** Boxplots of compounds with significant difference in the 6 h pool after intake of milk (A), cheese (B), and soy drink (C) assessed by NMR.

with pairwise comparison. As a result, 31 signals were statistically different with  $p < 0.05$  and among these, 22 had a  $p$ -value lower than 0.01. Eight signals appeared to relate to candidate biomarkers for cheese intake, 13 for milk and 14 for soy drink. These features are listed in Table S-5. In Figure 4, boxplots of the compounds with significant difference in the 6 h pool are shown.

The same analysis was applied by using the 1 by 1 comparison approach. A total of 58 variables were emerging by multivariate approach as discriminating among treatments, where further 2 variables were found with respect to the previous approach, although unassigned. However, these two extra variables were found with  $p > 0.05$  to the Kruskal–Wallis test, thus irrelevant as biomarker candidates.



The 24 h urine pool was then analyzed, in order to verify whether additional candidate markers, with a slower kinetics, could be identified. The new data set consisted of a total of 32 spectra, including 10 spectra for the milk intervention, 11 spectra for the soy drink intervention and 11 spectra for the cheese intervention. One sample for milk intervention was missing in one of the subjects, thus precluding the investigation with the inclusion of that subject. In this case, OPLS-DA calculated had  $Q^2$  values of, respectively, 0.732 (milk vs cheese/soy drink), 0.707 (cheese vs milk/soy drink), and 0.757 (soy drink vs milk/cheese), proving again a good predictability (Figure S-17). Compared to the 6 h pool, the lower  $Q^2$  values of the 24 h pool indicate that the most discriminating metabolites are metabolized within the first 6 h of the ingestion of the food products. The 60 most discriminating signals obtained from OPLS-DA were then analyzed by univariate statistical analysis (Kruskal–Wallis test with pairwise comparison), and 26 metabolites were identified as statistically different with  $p < 0.05$ . Also in this case, results from the two OPLS-DA approaches were not very different: 61 signals were identified with the 1 by 2 approach and 60 with the 1 by 1 approach. It is worth noting here that 58 signals were common to both sets, while the 5 signals differing between the two methods, i.e., 2-aminobutyrate, *N*-phenylacetyl glycine, U28 (1 by 2), and acetate and creatine (1 by 1), were not significant markers for specific foods according to the Kruskal–Wallis test. Among these metabolites, 17 had a  $p$ -value lower than 0.01. These signals mostly confirmed the finding of the 6 h urine pool, though the  $p$ -values were less significant as most candidate markers had fast excretion kinetics. These signals are also listed in Table S-5, allowing a comparison of the 6 h and 24 h pools.

**Evaluation of Candidate Markers.** Candidate biomarkers for milk intake, from both the 6 h and the 24 h pool, mainly corresponded to the typical sugars contained in milk, i.e. lactose and galactose. Not all signals originating from lactose and galactose were discriminative in both pools, evidently because of their kinetic of excretion but also because of a possible overlap with signals from the standardized meals taken 6 h after ingestion of the test meals. Interestingly a signal, which was assigned to galactonate was significantly increased in both pools only after milk intake. Also hippuric acid increased after milk ingestion, although this increase was statistically significant only for one of its signals falling in the sugar region of the spectrum, the other signals probably being affected by overlap with signals belonging to other metabolites. A singlet between 2.05 and 2.07 ppm, which could be assigned to an *N*-acetyl glycoprotein,<sup>30</sup> increased in four samples after milk ingestion. Compared to soy drink, many signals in this region were statistically increased in the 6 h pool after ingestion of the dairy products (data not shown), although only the singlet showed a statistically significant increase after milk consumption. The signal from allantoin at 5.39 ppm is also significantly increased after milk intake. A clear singlet is also increased in the 6 h urine pool after milk intake at 1.218 ppm. In this region, signals from the aliphatic moieties of molecules such as amino acids and fatty acids are observed. Dimethylamine, with a signal at 2.72 ppm is present in higher concentrations in the 6 h urine pool after consumption of dairy products. However, it is also present in high concentrations after soy drink intake. Two unassigned singlets, at 1.218 and 3.36 ppm were statistically discriminative for milk intake, although the first one only for the 6 h pool, while the second one for both pools.

Regarding cheese intake, the main candidate markers found were amino acids and organic acids. Among the amino acids, alanine was the most significant marker being at higher concentration after cheese consumption. Proline was also statistically increased after cheese intake compared to the other two products in the 6 h pool. The main organic acid showing the potential of being biomarker for cheese consumption was pyroglutamic acid. The assigned metabolites significantly increased after soy drink intake were pyridoxine, trigonelline and trans-aconitic acid. Finally, unassigned signals characteristic of aromatic molecules (8.5–7 ppm) were discriminative for soy drink intake.

**Intersubject Variability and Occurrence in Test Foods of Specific Candidate FIBs.** The intersubject variability of the identified FIBs defined as specific in this study is illustrated in Table 1 in the form of the CVs of the candidate FIBs within the

**Table 1. Intersubject Variability and Occurrence in Test Product of Specific Candidate Markers within This Study<sup>a</sup>**

Compound	CV % (intersubject)	CV % (QC)	CV(intersubject)/CV(QC)	Metabolite detected in food matrix
GC-MS				
Galactose	61	4	13.8	Milk
3-Phenyllactic acid	109	48	2.3	Cheese
D-Pinitol	29	4	7.0	Soy drink
Maltol	41	10	3.9	Soy drink
NMR				
Galactose	36	7	5	
<i>N</i> -Acetyl Glycoprotein	10	1	6.8	Milk
Pyridoxine	26	2	14.1	(overlap with other signals)

<sup>a</sup>The coefficients of variation (CVs) are shown for the markers measured by GC-MS (upper panel) and NMR (lower panel).

6 h pool. Intersubject variability has only been listed for identified compounds and for which the intersubject variability was higher than the technical variability obtained by replicate analysis of QC samples. Lactose (GC-MS, NMR) was thus excluded from the table as its technical variability with the current untargeted methods was higher than the intersubject variability; also galactonate (selected by NMR) has been excluded for the same reason. Table 1 also shows whether the candidate FIBs were present in their respective test products.

## DISCUSSION

### Candidate FIBs for Milk

Various metabolites derived from lactose were highly discriminating for milk intake. In particular, galactose and lactose were identified as specific candidate FIB for milk by both NMR and GC-MS methods while galactonate was detected by NMR. Even though galactitol and galactono-1,5-lactone were detected by GC-MS in urine after cheese and soy drink intake, these molecules might also be interesting markers of regular milk intake. Indeed, although they are detected under fasting conditions, they are clearly increased after milk intake. The presence of these compounds in urine has already been shown, including specifically lactose,<sup>31</sup> galactose,<sup>31</sup> galactonate,<sup>32</sup> galactitol,<sup>33</sup> and galactonolactone.<sup>34</sup> The presence of the last two compounds in urine was related to galactosemia, a

condition of galactokinase deficiency. Of note, in a very recent study, galactonate has been proposed as biomarker for dairy products intake.<sup>35</sup> Based on our study, we conclude that galactonate should not be considered as marker of dairy intake in general, as this metabolite was observed after milk but not cheese intake in our study.

It is not surprising that most discriminating markers for milk intake relate to lactose metabolism. During this test the volunteers received an unusually high amount of lactose in a single intake. Intestinal lactase is likely not able to efficiently metabolize such excessive amounts of lactose and non-hydrolyzed lactose was thus partially absorbed and excreted intact in urine. Part of the galactose molecule, hydrolyzed from lactose, would be expected to be metabolized in the Leloir pathway, which enzymatically converts it into UDP-glucose in the liver. This pathway depends upon the action of four enzymes, galactokinase being the rate-limiting enzyme.<sup>36</sup> If galactose is not metabolized efficiently, such as in case of excess galactose, the sugar will be catabolized by alternate pathways either to be converted to galactitol by NADPH-dependent aldose reductase or to galactonate catalyzed by galactose dehydrogenase.<sup>37</sup> The first intermediate product in the catabolism from galactose to galactonate is its lactone form, which occurs as 1,4- or 1,5-lactone in solution.<sup>38,39</sup> Therefore, at the amounts ingested in this study, our data indicates that galactose is metabolized via a range of biochemical processes including (i) metabolism by the Leloir pathway, (ii) reduction to galactitol, (iii) oxidation to galactonolactone (identified by GC-MS) and, further, to galactonate (identified by NMR) or (iv) excretion in nonmetabolized form. Whether these compounds would still be formed if the ingested amounts of lactose were lower remains uncertain. However, lower blood and urinary galactose concentrations would certainly be expected as galactose would be more completely metabolized by these pathways. In addition, as the formation of these compounds is based on enzymatic catalysis, it is highly likely that their metabolic fate is prone to high interindividual variation, a phenomena that is highlighted by the distinct CV measured in our study for galactose.

Looking at these candidate FIBs for milk intake from a broader point of view, it should be noted that these molecules are ultimately not fully specific for milk intake as, for instance, yogurt, sour milk and cream also contain lactose, although admittedly at lower levels than in milk. In addition, galactose in free form is also found in fruits and vegetables, although at levels lower than 40 mg/100g.<sup>40</sup> These foods might not confound the intake of dairy products since they are around 50 times lower than in milk (after hydrolysis of lactose). Consequently, although free galactose and its catabolized forms might be promising candidate FIBs for milk intake and other lactose-containing dairy products, their dose–response still needs to be investigated to precisely apply the marker as a FIB.

Allantoin was also demonstrated by NMR to be significantly increased after the intake of milk. This metabolite is the end product of purine metabolism in ruminants and is also found in cow's milk.<sup>41</sup> In humans, uric acid, and not allantoin, is the end product of purine metabolism.<sup>42</sup> Consequently, this metabolite might indeed quantitatively differentiate high milk consumers from nonconsumers. Hippurate, a milk marker of grazing grassland pasture for cows,<sup>43</sup> was also found in higher concentrations after milk intake, although only one of its numerous spectral signals were discriminant of milk con-

sumption. Hippurate was recently proposed as a marker of cheese intake by Zheng et al. (2015)<sup>19</sup> but in our work this metabolite was not increased after cheese intake. The production of hippurate is, however, dependent on the gut microbiota activity and Bertram et al. (2007)<sup>17</sup> reported decreased levels of hippurate after a milk intervention. Clarifying the potential of this metabolite as a FIB for the intake of dairy products would thus request an analysis of the composition of the microbiota. Another interesting signal that was specifically related to milk consumption was identified as N-acetyl glycoprotein (NAG). This molecule has been reported to discriminate lactovegetarian from omnivores.<sup>30</sup> However, a study by Van Dorsten et al. (2006),<sup>44</sup> linked it to the consumption of black and green tea, in a study on the consumption of caffeine. Thus, it is not clear whether this molecule can be exclusively linked to milk intake. Another metabolite that appeared to be a milk biomarker, based on NMR, was dimethylamine. Pairwise comparison suggested that this metabolite could be a biomarker of dairy intake (Table S-5); however after a visual inspection of the spectra, it was noted that for some samples from the soy drink intervention there were high levels of this metabolite. For this reason it is not possible to designate this metabolite the role of dairy intake biomarker.

#### Candidate FIBs for Cheese

FIBs to discriminate cheese and milk intake may be expected to be of microbial origin or formed during the process of cheese production and ripening. The amino acid derivative, 3-phenyllactic acid was the only specific FIB for cheese in this study. 3-phenyllactic acid derives from phenylalanine metabolism and its urinary levels are generally low.<sup>45</sup> This molecule is produced from a wide range of species of lactic acid bacteria (LAB).<sup>46</sup> Thus, its occurrence in urine after food intake may not only be attributed to the intake of cheese as a wide range of foods undergoes fermentation by lactic acid bacteria including nondairy products such as sourdough.<sup>47</sup> In addition, 3-phenyllactic acid is also used as biological preservative in the food industry due to its antimicrobial activity<sup>48</sup> and is commonly found in honey.<sup>49</sup> 3-Phenyllactic acid was detected in the cheese tested in this study (Le Gruyère AOP). To our knowledge differences in 3-phenyllactic acid concentrations in various types of cheese have not yet been studied so that it cannot be concluded that this compound would occur in urine after the intake of all types of cheese, although its high potential as FIB remains.

Due to the well-described proteolytic activity of LAB during cheese ripening, changes in the postprandial concentration of free amino acids are expected. In addition to alanine, which was identified as a nonspecific candidate marker of cheese intake by both analytical methods, proline was significantly increased after cheese intake in the NMR data. This amino acid is a typically found in Swiss cheese, and its concentration increases during ripening strongly impacting on the flavor of Gruyère.<sup>50</sup> Pedersen et al. found higher concentrations of proline in human serum after the consumption of a fermented yogurt product.<sup>18</sup> However, as amino acids are ubiquitously present in many foods and released during digestion, they cannot be regarded as FIB for cheese intake. Furthermore, it is unclear whether these two amino acids were increased in urine due to their higher bioavailability in the cheese matrix or simply because protein content was highest in cheese.

Another amino acid that was significantly increased after cheese intake was pyroglutamate (identified by NMR in urine samples). A trend for an increase in this metabolite was also shown by GC-MS. Pyroglutamate is present in cheese and has been proposed as a marker of cheese ripening.<sup>51,52</sup> Indeed, Ochi et al. reported pyroglutamic acid to contribute to the “rich flavour” of ripened cheeses.<sup>51</sup> The production of this metabolite is highly dependent on the starter bacteria used to ferment milk.<sup>52</sup> Pyroglutamate identified in the urine samples might thus directly originate from the cheese and be excreted nonmetabolized. It could thus be considered as a potential biomarker of cheese intake. As GC-MS only found a slight increase in pyroglutamate concentration after cheese intake, more studies are necessary to confirm its validity as a cheese biomarker.

### Candidate FIBs for the Soy Drink

The control product consisted of 90% soy drink and 10% added soy based vegetable cream. The markers found for this product thus refer to this mixture but most likely derive from the soy drink. D-Pinitol and maltol were both specific to soy drink intake. D-Pinitol is a cyclic polyol, a family of compounds abundant in soybeans. D-Pinitol is also found in other edible plants from the *Fabaceae* family such as clover, locust tree and carob,<sup>53</sup> although these products are consumed at much lower quantities than soy products. The second compound, maltol, is used as flavor enhancer in bakery products but is known to occur in foods such as soybeans, coffee and chicory.<sup>54</sup> In our study, both compounds were highly discriminating of soy drink intake. Urinary sucrose was also discriminative for soy drink intake but, since it is ubiquitously present in food and beverages at much higher concentration, it should be excluded from a list of candidate markers for soy drink intake. The same argument is valid for the inositol-like compound identified in our study since inositol is ubiquitously present in fruits, beans, grains and nuts.<sup>55</sup>

Catechol, one of the late urinary markers, may derive from the microbial degradation of aromatic compounds.<sup>56,57</sup> It can only be assumed that catechol found after soy drink intake derives from vanillic acid, as this hydrobenzoic acid is present in soy.<sup>58</sup> However, the concentration of catechol in urine has previously been shown to decrease when the intake of plant-derived foods was reduced.<sup>59</sup> Thus, the presence of this compound in urine is most certainly not restricted to soy drink intake. Guaiacol showed exactly the same excretion pattern as catechol. It is a known intermediate compound in the microbial conversion of vanillic acid to catechol.<sup>57</sup> Thus, the appearance of both compounds in the late postprandial phase can most likely be attributed to their microbial formation in the gut.

The main candidate FIBs identified by NMR for soy drink intake were trigonelline, pyridoxine and trans-aconitate. Trigonelline is present in soybeans.<sup>60</sup> However, this metabolite is also a typical compound that accumulates in legumes and coffee seeds and that has been already linked to coffee intake.<sup>61</sup> Pyridoxine (vitamin B6) also showed increased urinary concentrations after soy drink intake. Soy products are a good source of this vitamin although soy-derived products may have significantly lower concentrations compared to soybeans.<sup>62</sup> Trans-aconitate has been found by NMR in the urine of soy drink consumers. This planted-derived organic acid is not only present in soybean but also in wheat<sup>63</sup> and, thus, cannot be considered as a fully specific soy drink marker. Acetoacetate was also discriminative of soy drink intake using

the Kruskal–Wallis analysis of the NMR spectra. The concentration of this molecule increased in samples after the soy drink and the cheese intervention, although at respectively lower levels in the latter. Acetoacetate concentration was reportedly higher in plasma after a soy intervention<sup>64</sup> compared to a control diet. The authors correlated this increase to a shift in lipid metabolism. In spite of this, acetoacetate cannot be completely considered as biomarker for a specific food intake, due to the fact that it appears in both cheese and soy drink. Moreover, the relevance of this molecule as FIB is questionable because not all its signals were discriminant by univariate statistical analysis. Finally, a series of unassigned NMR signals and GC-MS peaks could be ascribed to isoflavone-like molecules. In particular, a doublet found at 6.86 ppm, in the NMR spectra was tentatively assigned to daidzein, one of the most studied isoflavones found in soy products. However, additional work is necessary to robustly assign these compounds to well-defined isoflavone metabolites and to propose them as FIBs for soy drink intake.

### Intersubject Variability and Occurrence of FIBs in Products

All of the level 1 identified specific candidate FIBs (lactose, galactose, 3-phenyllactic acid, D-pinitol and maltol) were already detected by GC-MS in the test product for which they were assigned as markers. Free galactose was found in milk but given the large quantity of urinary galactose, it is surmised that much of the metabolite that was detected in urine must have been formed during hydrolysis of lactose at the intestinal brush border. Thus, four out of five specific candidate FIBs found by GC-MS (lactose, 3-phenyllactic acid, D-pinitol and maltol) were most likely nonmetabolized or not fully metabolized compounds, thus, being detected after the intake of specific food. In the case of NMR, galactose was not detected in the product. For mono- and disaccharides, anomeric signals are selected for identification and quantification taking advantage of their diagnostic power. However, by inspecting the NMR spectrum, the galactose anomeric signal in milk was faint and partially overlapped with other minor saccharides' signals so that it was decided not to consider it as unambiguously assigned.

The intersubject variability of the specific compounds was within or even below reported ranges for urinary metabolites.<sup>31</sup> As shown by Bouatra et al. (2013)<sup>31</sup> most metabolites in urine varied by about 50% while some vary by up to 350%. The CVs measured for each metabolite in the urinary samples originate primarily from two sources, interindividual variability and analytical variability. Intersubject variability can be due to differences in genetics, lifestyle including dietary habits, and the composition of the gut microbiota. These factors not only change the metabolism of nutrients quantitatively but also qualitatively by modulating the relative use of different pathways taken during food digestion and metabolism.<sup>65</sup> Taken together, our data indicate that the urinary concentrations of the specific FIBs within this study are specific to each individual. The intersubject variability in the metabolic processing of these candidate FIBs should thus be taken into consideration for their validation. In addition, targeted analyses with improved CV would be needed, in particular with metabolites with high CV, such as lactose.

### Strengths and Limitations of the Study

Our study highlights the synergistic power of a multiplatform (GC-MS and NMR) nutrimentalomic approach for either confirming the results for metabolites detected by both



techniques or for identifying additional candidate FIBs that are better evaluated on either one of the platforms. Moreover, the different characteristics of the two techniques (i.e., high sensitivity for GC-MS, high precision and repetitiveness for NMR) combined with the identification of the same molecules by both methods supported the robustness of the results.

Gender as well as BMI have an effect on the human metabolome.<sup>66,67</sup> However, as FIBs should be independent from gender and BMI, our study deliberately included both genders with a broad BMI range. Moreover, the male subjects had higher BMIs ( $23.2 \pm 1.3 \text{ kg/m}^2$ ) than female subjects ( $20.9 \pm 1.1 \text{ kg/m}^2$ ), which prevented us from separating the effects of BMI from the effects of gender.

GC-MS analysis compared the urine samples after intake of each food to the grouped samples after intake of the two other foods, i.e. in a 1 by 2 pairwise design. Besides the 1 by 2 pairwise design, the NMR analysis compared each possible combination of foods also in a 1 by 1 pairwise design, resulting with any different significant biomarker between the two methods.

A large quantity of each food that was tested in order to favor the identification of candidate FIBs, which could subsequently be validated in real life conditions. One limitation of this approach for milk is that the candidate FIBs could arise from the excess intake of lactose and thus an overload of the pathways related to the metabolism of this sugar, in particular galactose metabolism. Therefore, further studies are needed to demonstrate that lactose-related candidate FIBs can also be found in urine when less milk is consumed. In addition, the use of lactose and lactose-derived metabolites as exposure markers for milk is faced with the uncertainties related to the large variability in the content of lactose and galactose present in other dairy products. Consequently, these markers might finally be FIBs for the intake of lactose-containing products rather than only milk.

## CONCLUSION

None of the urinary biomarkers for milk and cheese intake previously reported have been confirmed by this study. In this study, some urinary candidate FIBs were specific for one food compared to the other two foods and showed clear time-responses fulfilling some criteria of an ideal FIB. However, these molecules will most certainly not be specific in comparison to a broader range of foods. In addition, the relevance of these compounds as FIBs for milk, cheese and soy based foods should be assessed in observational studies under noncontrolled conditions and within a broader dietary context. Further, dose–response studies should be performed to further validate the compounds identified in this study for their use as FIBs. Such studies will require a quantitative assessment of these molecules within the urine matrix using targeted analysis. In this study, markers for milk and cheese intake were excreted within 6 h. The pooling of urine up to 6 h would thus be sufficient to detect these dairy FIBs in contrast to the plant-based control, for which late urinary markers also appeared. Finally, combining compounds based on a structured definition of the foods of interest, e.g. combining markers of milk with markers of fermentation to identify markers of cheese intake, might be a promising strategy that would also confirm adherence of consumers to the broader dietary guidelines that often relate to food groups rather than single foods.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.7b00319.

Test product composition; study population characteristics; internal numbering system of compounds analyzed by GC-MS, retention index (RI), spectral data used for targeted evaluation, identification levels of features, and corresponding coefficient of variations (CV) in 6 h pools, 24 h pools, and/or interval samples; ID from deconvolution data table and respective VIP values from OPLS-DA model, and *p*-values obtained after targeted evaluation of features in respective pool samples as well as internal numbering system name; multivariate analysis of 6 h pools; score plot, loading plot with most discriminating features in red, permutation test ( $n = 999$ ) and model parameters of OPLS-DA model based on 6 h pool samples; multivariate analysis of 24 h pools (milk and soy drink) and interval samples (cheese); score plot, loading plot, permutation test, and model parameters of OPLS-DA model; level 1 identifications of D-galactose, D-lactose, 3-phenyllactic acid, D-pinitol, maltol, galactitol, L-alanine, sucrose, catechol, guaiacol, pyroglutamic acid, 4-methylcatechol, and compound M09 with reference compounds; boxplots of unknown compounds; list of discriminative signals found for each product by NMR; OPLS-DA score-plots; and list of metabolites found by NMR spectroscopy. (PDF)

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [rene.badertscher@agroscope.admin.ch](mailto:rene.badertscher@agroscope.admin.ch). Phone: 0041 58 46 381 82.

### ORCID

René Badertscher: 0000-0001-9318-6470

### Funding

The study was funded by the Swiss National Science Foundation (40HD40\_160618) in the frame of the national research program “Healthy nutrition and sustainable food protection (NRP69)” and the HDHL Joint Programming Initiative “The Food Biomarkers Alliance (FoodBALL)”.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank Doreen Gille, Reto Portmann, Alexandra Schmid, and Barbara Walther for their assistance in designing and in conducting the clinical trial, Edith Beutler for the preparation of standardised meals, Lotti Egger for the food analyses and Pascal Fuchsmann for refractive index analyses (all Agroscope), as well as Ismael Turk MD for the organization and conduction of the clinical trial, and Marie-Jeanne Voirol and Alexia Spoerl (all CHUV) for contributions to laboratory work and sample collection.

## ABBREVIATIONS

FIB, food intake biomarker; FoodBALL, Food Biomarker Alliance; GC, gas chromatography; GMD, Golm Metabolome Database; HMDB, Human Metabolome Database; IStd,



internal standard; LAB, lactic acid bacteria; MS, mass spectrometry; NMR, nuclear magnetic resonance; OPLS-DA, Orthogonal Projections to Latent Structures Discriminant Analysis; PCA, principal component analysis; QC, quality control; TSP, trimethylsilyl propionate; RI, retention index; RT, retention time; VIP, variable importance of projection

## REFERENCES

- (1) FOOTBALL Project Overview. <http://foodmetabolome.org/> (July 01 2017).
- (2) Jenab, M.; Slimani, N.; Bictash, M.; Ferrari, P.; Bingham, S. A. Biomarkers in nutritional epidemiology: applications, needs and new horizons. *Hum. Genet.* **2009**, *125* (5–6), 507–25.
- (3) Scalbert, A.; Brennan, L.; Manach, C.; Andres-Lacueva, C.; Dragsted, L. O.; Draper, J.; Rappaport, S. M.; van der Hooft, J. J.; Wishart, D. S. The food metabolome: a window over dietary exposure. *Am. J. Clin. Nutr.* **2014**, *99* (6), 1286–308.
- (4) Trimigno, A.; Picone, G.; Capozzi, F. A 1H NMR-Based Metabolomics Approach on Dietary Biomarker Research in Human Urine. *Magnetic Resonance in Food Science: Defining Food by Magnetic Resonance* **2015**, 143.
- (5) Bordoni, A.; Capozzi, F. Foodomics for healthy nutrition. *Curr. Opin. Clin. Nutr. Metab. Care* **2014**, *17* (5), 418–24.
- (6) Heinzmann, S. S.; Brown, I. J.; Chan, Q.; Bictash, M.; Dumas, M. E.; Kochhar, S.; Stamler, J.; Holmes, E.; Elliott, P.; Nicholson, J. K. Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. *Am. J. Clin. Nutr.* **2010**, *92* (2), 436–43.
- (7) Lloyd, A. J.; Beckmann, M.; Fave, G.; Mathers, J. C.; Draper, J. Proline betaine and its biotransformation products in fasting urine samples are potential biomarkers of habitual citrus fruit consumption. *Br. J. Nutr.* **2011**, *106* (6), 812–24.
- (8) Vazquez-Fresno, R.; Llorach, R.; Perera, A.; Mandal, R.; Feliz, M.; Tinahones, F. J.; Wishart, D. S.; Andres-Lacueva, C. Clinical phenotype clustering in cardiovascular risk patients for the identification of responsive metabolotypes after red wine polyphenol intake. *J. Nutr. Biochem.* **2016**, *28*, 114–20.
- (9) Kristensen, M.; Engelsen, S. B.; Dragsted, L. O. LC–MS metabolomics top-down approach reveals new exposure and effect biomarkers of apple and apple-pectin intake. *Metabolomics* **2012**, *8* (1), 64–73.
- (10) Cuparencu, C. S.; Andersen, M.-B. S.; Gürdeniz, G.; Schou, S. S.; Mortensen, M. W.; Raben, A.; Astrup, A.; Dragsted, L. O. Identification of urinary biomarkers after consumption of sea buckthorn and strawberry, by untargeted LC–MS metabolomics: a meal study in adult men. *Metabolomics* **2016**, *12* (2), 31.
- (11) Pujos-Guillot, E.; Hubert, J.; Martin, J. F.; Lyan, B.; Quintana, M.; Claude, S.; Chabanas, B.; Rothwell, J. A.; Bennetau-Pelissero, C.; Scalbert, A.; Comte, B.; Hercberg, S.; Morand, C.; Galan, P.; Manach, C. Mass spectrometry-based metabolomics for the discovery of biomarkers of fruit and vegetable intake: citrus fruit as a case study. *J. Proteome Res.* **2013**, *12* (4), 1645–59.
- (12) O’Gorman, A.; Brennan, L. The role of metabolomics in determination of new dietary biomarkers. *Proc. Nutr. Soc.* **2017**, 1–8.
- (13) Golley, R. K.; Hendrie, G. A. Evaluation of the relative concentration of serum fatty acids C14:0, C15:0 and C17:0 as markers of children’s dairy fat intake. *Ann. Nutr. Metab.* **2014**, *65* (4), 310–6.
- (14) Meikle, P. J.; Barlow, C. K.; Mellett, N. A.; Mundra, P. A.; Bonham, M. P.; Larsen, A.; Cameron-Smith, D.; Sinclair, A.; Nestel, P. J.; Wong, G. Postprandial Plasma Phospholipids in Men Are Influenced by the Source of Dietary Fat. *J. Nutr.* **2015**, *145* (9), 2012–8.
- (15) Abdullah, M. M. H.; Cyr, A.; Lépine, M. C.; Labonté, M. È.; Couture, P.; Jones, P. J. H.; Lamarche, B. Recommended dairy product intake modulates circulating fatty acid profile in healthy adults: A multi-centre cross-over study. *Br. J. Nutr.* **2015**, *113* (3), 435–444.
- (16) Lankinen, M.; Schwab, U. Biomarkers of dairy fat. *Am. J. Clin. Nutr.* **2015**, *101* (5), 1101–2.
- (17) Bertram, H. C.; Hoppe, C.; Petersen, B. O.; Dues, J. O.; Molgaard, C.; Michaelsen, K. F. An NMR-based metabolomic investigation on effects of milk and meat protein diets given to 8-year-old boys. *Br. J. Nutr.* **2007**, *97* (4), 758–763.
- (18) Pedersen, S. M. M.; Nebel, C.; Nielsen, N. C.; Andersen, H. J.; Olsson, J.; Simren, M.; Ohman, L.; Svensson, U.; Bertram, H. C.; Malmendal, A. A GC-MS-based metabolomic investigation of blood serum from irritable bowel syndrome patients undergoing intervention with acidified milk products. *Eur. Food Res. Technol.* **2011**, *233* (6), 1013–1021.
- (19) Zheng, H.; Yde, C. C.; Clausen, M. R.; Kristensen, M.; Lorenzen, J.; Astrup, A.; Bertram, H. C. Metabolomics Investigation To Shed Light on Cheese as a Possible Piece in the French Paradox Puzzle. *J. Agric. Food Chem.* **2015**, *63* (10), 2830–2839.
- (20) Hjerpested, J. B.; Ritz, C.; Schou, S. S.; Tholstrup, T.; Dragsted, L. O. Effect of cheese and butter intake on metabolites in urine using an untargeted metabolomics approach. *Metabolomics* **2014**, *10* (6), 1176–1185.
- (21) Edmands, W. M.; Ferrari, P.; Scalbert, A. Normalization to specific gravity prior to analysis improves information recovery from high resolution mass spectrometry metabolomic profiles of human urine. *Anal. Chem.* **2014**, *86* (21), 10925–31.
- (22) Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Kell, D. B.; Goodacre, R. Human Serum Metabolome, C., Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **2011**, *6* (7), 1060–83.
- (23) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M. Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **2012**, *30* (10), 918–20.
- (24) Domingo-Almenara, X.; Brezmes, J.; Vinaixa, M.; Samino, S.; Ramirez, N.; Ramon-Krauel, M.; Lerin, C.; Diaz, M.; Ibanez, L.; Correig, X.; Perera-Lluna, A.; Yanes, O. eRah: A Computational Tool Integrating Spectral Deconvolution and Alignment with Quantification and Identification of Metabolites in GC/MS-Based Metabolomics. *Anal. Chem.* **2016**, *88* (19), 9821–9829.
- (25) Kopka, J.; Schauer, N.; Krueger, S.; Birkemeyer, C.; Usadel, B.; Bergmuller, E.; Dormann, P.; Weckwerth, W.; Gibon, Y.; Stitt, M.; Willmitzer, L.; Fernie, A. R.; Steinhauser, D. GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics* **2005**, *21* (8), 1635–8.
- (26) Triba, M. N.; Le Moyec, L.; Amathieu, R.; Goossens, C.; Bouchemal, N.; Nahon, P.; Rutledge, D. N.; Savarin, P. PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. *Mol. BioSyst.* **2015**, *11* (1), 13–9.
- (27) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A. N.; Lindon, J. C.; Marriott, P.; Nicholls, A. W.; Reilly, M. D.; Thaden, J. J.; Viant, M. R. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **2007**, *3* (3), 211–221.
- (28) Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabolomics. *Anal. Chem.* **2006**, *78* (13), 4281–4290.
- (29) Dunn, W. B.; Lin, W.; Broadhurst, D.; Begley, P.; Brown, M.; Zelena, E.; Vaughan, A. A.; Halsall, A.; Harding, N.; Knowles, J. D.; Francis-McIntyre, S.; Tseng, A.; Ellis, D. I.; O’Hagan, S.; Aarons, G.;

Benjamin, B.; Chew-Graham, S.; Moseley, C.; Potter, P.; Winder, C. L.; Potts, C.; Thornton, P.; McWhirter, C.; Zubair, M.; Pan, M.; Burns, A.; Cruickshank, J. K.; Jayson, G. C.; Purandare, N.; Wu, F. C.; Finn, J. D.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Goodacre, R.; Kell, D. B. Molecular phenotyping of a UK population: defining the human serum metabolome. *Metabolomics* **2015**, *11*, 9–26.

(30) Xu, J.; Yang, S.; Cai, S.; Dong, J.; Li, X.; Chen, Z. Identification of biochemical changes in lactovegetarian urine using 1H NMR spectroscopy and pattern recognition. *Anal. Bioanal. Chem.* **2010**, *396* (4), 1451–63.

(31) Bouatra, S.; Aziat, F.; Mandal, R.; Guo, A. C.; Wilson, M. R.; Knox, C.; Bjorn Dahl, T. C.; Krishnamurthy, R.; Saleem, F.; Liu, P.; Dame, Z. T.; Poelzer, J.; Huynh, J.; Yallou, F. S.; Psychogios, N.; Dong, E.; Bogumil, R.; Roehring, C.; Wishart, D. S. The human urine metabolome. *PLoS One* **2013**, *8* (9), e73076.

(32) Wehrli, S. L.; Berry, G. T.; Palmieri, M.; Mazur, A.; Elsas, L., 3rd; Segal, S. Urinary galactonate in patients with galactosemia: quantitation by nuclear magnetic resonance spectroscopy. *Pediatr. Res.* **1997**, *42* (6), 855–61.

(33) Jakobs, C.; Schweitzer, S.; Dorland, B. Galactitol in galactosemia. *Eur. J. Pediatr.* **1995**, *154* (Suppl2), S50–2.

(34) Rakotomanga, S.; Baillet, A.; Pellerin, F.; Baylocq-Ferrier, D. Simultaneous determination of gluconolactone, galactonolactone and galactitol in urine by reversed-phase liquid chromatography: application to galactosemia. *J. Chromatogr., Biomed. Appl.* **1991**, *570* (2), 277–84.

(35) Playdon, M. C.; Moore, S. C.; Derkach, A.; Reedy, J.; Subar, A. F.; Sampson, J. N.; Albanes, D.; Gu, F.; Kontto, J.; Lassale, C.; Liao, L. M.; Mannisto, S.; Mondul, A. M.; Weinstein, S. J.; Irwin, M. L.; Mayne, S. T.; Stolzenberg-Solomon, R. Identifying biomarkers of dietary patterns by using metabolomics. *Am. J. Clin. Nutr.* **2017**, *105* (2), 450–465.

(36) Keiding, S.; Johansen, S.; Winkler, K.; Tonnesen, K.; Tygstrup, N. Michaelis-Menten kinetics of galactose elimination by the isolated perfused pig liver. *American journal of physiology* **1976**, *230* (5), 1302–13.

(37) Coelho, A. I.; Berry, G. T.; Rubio-Gozalbo, M. E. Galactose metabolism and health. *Curr. Opin. Clin. Nutr. Metab. Care* **2015**, *18* (4), 422–7.

(38) Bierenstiel, M.; Schlaf, M.  $\delta$ -Galactonolactone: Synthesis, Isolation, and Comparative Structure and Stability Analysis of an Elusive Sugar Derivative. *Eur. J. Org. Chem.* **2004**, *2004* (7), 1474–1481.

(39) Cuatrecasas, P.; Segal, S. Galactose conversion to D-xylulose: an alternate route of galactose metabolism. *Science* **1966**, *153* (3735), 549–51.

(40) Kim, H. O.; Hartnett, C.; Scaman, C. H. Free galactose content in selected fresh fruits and vegetables and soy beverages. *J. Agric. Food Chem.* **2007**, *55* (20), 8133–7.

(41) Ackroyd, H. On the Presence of Allantoin in Certain Foods. *Biochem. J.* **1911**, *5* (8–9), 400–6.

(42) Maiuolo, J.; Oppedisano, F.; Gratteri, S.; Muscoli, C.; Mollace, V. Regulation of uric acid metabolism and excretion. *Int. J. Cardiol.* **2016**, *213*, 8–14.

(43) Besle, J. M.; Viala, D.; Martin, B.; Pradel, P.; Meunier, B.; Berdague, J. L.; Fraisse, D.; Lamaison, J. L.; Coulon, J. B. Ultraviolet-absorbing compounds in milk are related to forage polyphenols. *J. Dairy Sci.* **2010**, *93* (7), 2846–56.

(44) Van Dorsten, F. A.; Daykin, C. A.; Mulder, T. P.; Van Duynhoven, J. P. Metabonomics approach to determine metabolic differences between green tea and black tea consumption. *J. Agric. Food Chem.* **2006**, *54* (18), 6929–38.

(45) Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J.; Liu, P.; Yallou, F.; Bjorn Dahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A. HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Res.* **2013**, *41* (Database issue), D801–7.

(46) Valerio, F.; Lavermicocca, P.; Pascale, M.; Visconti, A. Production of phenyllactic acid by lactic acid bacteria: an approach to the selection of strains contributing to food quality and preservation. *FEMS Microbiol. Lett.* **2004**, *233* (2), 289–95.

(47) Lavermicocca, P.; Valerio, F.; Evidente, A.; Lazzaroni, S.; Corsetti, A.; Gobetti, M. Purification and characterization of novel antifungal compounds from the sourdough *Lactobacillus plantarum* strain 21B. *Appl. Environ. Microbiol.* **2000**, *66* (9), 4084–90.

(48) Mu, W.; Yu, S.; Zhu, L.; Zhang, T.; Jiang, B. Recent research on 3-phenyllactic acid, a broad-spectrum antimicrobial compound. *Appl. Microbiol. Biotechnol.* **2012**, *95* (5), 1155–63.

(49) Tuberoso, C. I.; Bifulco, E.; Caboni, P.; Sarais, G.; Cottiglia, F.; Floris, I. Lumichrome and phenyllactic acid as chemical markers of thistle (*Galactites tomentosa* Moench) honey. *J. Agric. Food Chem.* **2011**, *59* (1), 364–9.

(50) Zerfidis, G. K.; Vafopoulou-Mastrogiannaki, A.; Litopoulou-Tzanetaki, E. Changes during ripening of commercial Gruyère cheese. *J. Dairy Sci.* **1984**, *67* (7), 1397–1405.

(51) Ochi, H.; Naito, H.; Iwatsuki, K.; Bamba, T.; Fukusaki, E. Metabolomics-based component profiling of hard and semi-hard natural cheeses with gas chromatography/time-of-flight-mass spectrometry, and its application to sensory predictive modeling. *J. Biosci Bioeng* **2012**, *113* (6), 751–8.

(52) Mucchetti, G.; Locci, F.; Gatti, M.; Neviani, E.; Addeo, F.; Dossena, A.; Marchelli, R. Pyroglutamic acid in cheese: presence, origin, and correlation with ripening time of Grana Padano cheese. *J. Dairy Sci.* **2000**, *83* (4), 659–65.

(53) Negishi, O.; Mun'im, A.; Negishi, Y. Content of methylated inositols in familiar edible plants. *J. Agric. Food Chem.* **2015**, *63* (10), 2683–8.

(54) Barroso, M. F.; Sales, M. G. F.; Almeida, S. A. A.; Vaz, M. C. V. F.; Delerue-Matos, C. Maltol Determination in Food by Microwave Assisted Extraction and Electrochemical Detection. *Journal of Food & Drug Analysis* **2008**, *16* (4), 30–36.

(55) Clements, R. S., Jr.; Darnell, B. Myo-inositol content of common foods: development of a high-myo-inositol diet. *Am. J. Clin. Nutr.* **1980**, *33* (9), 1954–67.

(56) Evans, W. C. The Microbiological Degradation of Aromatic Compounds. *J. Gen. Microbiol.* **1963**, *32*, 177–84.

(57) Alvarez-Rodriguez, M. L.; Belloch, C.; Villa, M.; Uruburu, F.; Larriba, G.; Coque, J. J. Degradation of vanillic acid and production of guaiacol by microorganisms isolated from cork samples. *FEMS Microbiol. Lett.* **2003**, *220* (1), 49–55.

(58) Xu, B.; Chang, S. K. Total phenolics, phenolic acids, isoflavones, and anthocyanins and antioxidant properties of yellow and black soybeans as affected by thermal processing. *J. Agric. Food Chem.* **2008**, *56* (16), 7165–75.

(59) Carmella, S. G.; La Voie, E. J.; Hecht, S. S. Quantitative analysis of catechol and 4-methylcatechol in human urine. *Food Chem. Toxicol.* **1982**, *20* (5), 587–90.

(60) Harrigan, G. G.; Skogerson, K.; MacIsaac, S.; Bickel, A.; Perez, T.; Li, X. Application of (1)h NMR profiling to assess seed metabolomic diversity. A case study on a soybean era population. *J. Agric. Food Chem.* **2015**, *63* (18), 4690–7.

(61) Madrid-Gambin, F.; Garcia-Aloy, M.; Vázquez-Fresno, R.; Vegas-Lozano, E.; de Villa Jubany, M. C. R.; Misawa, K.; Hase, T.; Shimotoyodome, A.; Andres-Lacueva, C. Impact of chlorogenic acids from coffee on urine metabolome in healthy human subjects. *Food Res. Int.* **2016**, *89* (Part 3), 1064–1070.

(62) Lebidzińska, A.; Szefer, P. Vitamins B in grain and cereal–grain food, soy-products and seeds. *Food Chem.* **2006**, *95* (1), 116–122.

(63) Yuhara, K.; Yonehara, H.; Hattori, T.; Kobayashi, K.; Kirimura, K. Enzymatic characterization and gene identification of aconitate isomerase, an enzyme involved in assimilation of trans-acetic acid, from *Pseudomonas* sp. WU-0701. *FEBS J.* **2015**, *282* (22), 4257–67.

(64) Solanky, K. S.; Bailey, N. J.; Beckwith-Hall, B. M.; Davis, A.; Bingham, S.; Holmes, E.; Nicholson, J. K.; Cassidy, A. Application of biofluid 1H nuclear magnetic resonance-based metabonomic techni-

ques for the analysis of the biochemical effects of dietary isoflavones on human plasma profile. *Anal. Biochem.* **2003**, *323* (2), 197–204.

(65) Flint, H. J.; Duncan, S. H.; Scott, K. P.; Louis, P. Links between diet, gut microbiota composition and gut metabolism. *Proc. Nutr. Soc.* **2015**, *74* (1), 13–22.

(66) Thevenot, E. A.; Roux, A.; Xu, Y.; Ezan, E.; Junot, C. Analysis of the Human Adult Urinary Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses. *J. Proteome Res.* **2015**, *14* (8), 3322–35.

(67) Moore, S. C.; Matthews, C. E.; Sampson, J. N.; Stolzenberg-Solomon, R. Z.; Zheng, W.; Cai, Q.; Tan, Y. T.; Chow, W. H.; Ji, B. T.; Liu, D. K.; Xiao, Q.; Boca, S. M.; Leitzmann, M. F.; Yang, G.; Xiang, Y. B.; Sinha, R.; Shu, X. O.; Cross, A. J. Human metabolic correlates of body mass index. *Metabolomics* **2014**, *10* (2), 259–269.