



Contents lists available at SciOpen

Food Science and Human Wellness

journal homepage: <https://www.sciopen.com/journal/2097-0765>

## Urinary Biomarkers of Soybean and Corn Intake Identified through Untargeted Metabolomics in a Randomized Cross-Over Trial

Huini Ding <sup>a</sup>, Chengyan Qi <sup>a</sup>, Hongmei Huang <sup>b</sup>, Jing Shen <sup>a</sup>, Yafei Chi <sup>c</sup>, Dajun Li <sup>a</sup>, Kaifeng Li <sup>d</sup>,  
Huan Peng <sup>e</sup>, Zhao Wang <sup>e</sup>, Yuandi Xi <sup>a\*</sup>

<sup>a</sup> Beijing Key Laboratory of environment and aging, School of Public Health, Capital Medical University, Beijing 100069, China.

<sup>b</sup> Department of Clinical Nutrition, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing 100045, China.

<sup>c</sup> Department of Experimental Animals, Capital Medical University, Beijing 100069, China.

<sup>d</sup> Feihe Research Institute, Heilongjiang Feihe Dairy Co., Ltd, Beijing 100016, China.

<sup>e</sup> Protein Science Key Laboratory of the Ministry of Education, School of Pharmaceutical Sciences, Tsinghua University, Beijing 100084, China.

**ABSTRACT:** Biomarkers of soybean and whole-grain cereal intake are insufficiently studied in Chinese populations. In this randomized cross-over trial, 24 participants received four different breakfast interventions during 8 weeks. Each intervention was preceded by a 5-day run-in period and followed by an 8-day washout period. Participants were randomly assigned one of four diets per intervention: Breakfast A (highest oat content: 75 g), Breakfast B (highest millet content: 75 g), Breakfast C (highest soy products content: 325 g), Breakfast D (highest corn content: 100 g). Urine samples collected at 1, 2, 3, 4, 6, 8, 12, and 24 hours after the meal were pooled for untargeted metabolomics. Univariate analysis, linear regression and random forest were used to identify differential metabolites. We found that 3 metabolites in Breakfasts A and B, and 14 metabolites in Breakfasts C and D differed in their relative concentration. Daidzein, genistein 5-O-glucuronide, genistein 4'-O-glucuronide, and glycitein 7-O-glucuronide exhibited the highest fold-change (FC) values (exceeding 25.0) and importance rankings following the 325 g soybean intervention, and demonstrated significant linear correlations with soybean intake ( $R^2=0.96-0.99$ ). Indole-3-acetic acid (IAA), indole-3-carbinol (I3C), and 1H-indole-3-carboxaldehyde (IAld) increased progressively from Breakfast C (no corn) to D (100 g corn), with FC of 0.8, 0.7 and 0.6. IAA, I3C and IAld were consistent with the trend of corn intake ( $R^2=0.91, 0.95, 0.58$ ). Future studies are needed to clarify plausibility and validity of these potential biomarkers for soybean and corn intakes.

**Keywords:** Soybean; Corn; Biomarkers; Metabolomics; Urinary metabolites

### 1. Introduction

Corn, millet, and oats are whole grain cereals that constitute nutritionally indispensable dietary components in China. They are rich in B vitamins, minerals, dietary fiber and other essential nutrients <sup>[1,2]</sup>. Soybeans and soy products serve as excellent protein sources <sup>[3]</sup> and are notably abundant in lysine. This is particularly valuable as lysine is the limiting amino acid in wheat and many other cereals, making soy an ideal nutritional complement to cereal-based diets <sup>[4]</sup>. Traditional dietary assessment methods such as food

\*Corresponding author  
sdykd2020@163.com

Received 13 October 2025  
Received in revised form 24 November 2025  
Accepted 9 February 2026

frequency questionnaires, are often subject to recall bias and measurement errors <sup>[5]</sup>. The application of metabolomics to identify characteristic food biomarkers in blood or urine <sup>[6]</sup> offers an objective approach to improve the accuracy of food intake assessment in observational studies <sup>[7]</sup> and to address compliance issues in dietary intervention research <sup>[8]</sup>. Furthermore, this methodology can enhance our understanding of the health benefits associated with foods, including cardiovascular protection <sup>[9]</sup>, blood glucose regulation and anti-inflammatory effects <sup>[10]</sup>. According to the Chinese Dietary Guidelines, the recommended daily intake of whole grains and mixed beans for adults is 50–150 g <sup>[11]</sup>. However, the Report on the *Nutrition and Chronic Disease Status of Chinese Residents (2020)*, issued by the National Health Commission of China, indicates that over 80% of Chinese adults fail to meet this target. Furthermore, across all age groups in China, the average daily intake of soybeans and soy products remains below the recommended dietary level of 15–25 g/day. Developing reliable biomarkers can provide unbiased evidence for changing population trends in the consumption of foods important for public health, and from this inform public health strategies.

Current research on dietary biomarkers for whole grains cereals and soy products has primarily focused on Western dietary patterns <sup>[12–14]</sup>, with few studies attempting to discover or validate biomarkers within Eastern diets. Although alkylresorcinols (ARs) and their metabolites have been validated as biomarkers of whole-grain wheat and rye intake <sup>[15–17]</sup>, they are not suitable for assessing intake of other grains such as millet, corn, or oats, as these do not contain ARs <sup>[18]</sup>. Given the high consumption of these grains in China, there is a clear need to identify and validate alternative biomarkers.

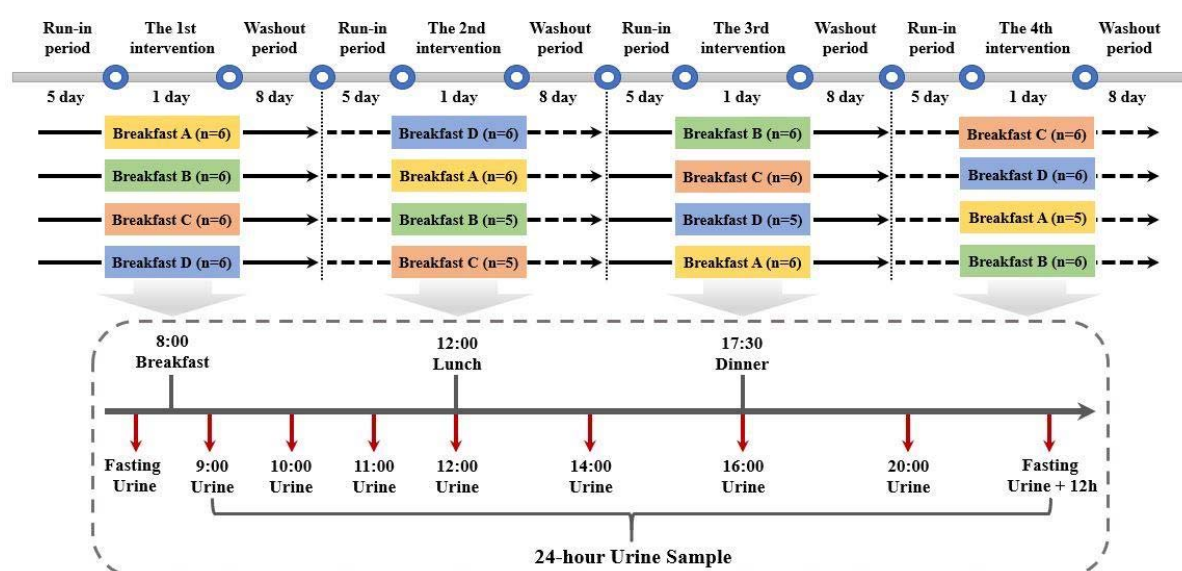
Regarding soy products, a randomized controlled trial reported increased urinary isoflavone levels (daidzein, glycitein, genistein and equol) after six months of soy flour intake <sup>[19]</sup>, while another study found a significant increase in plasma total isoflavones following two years of dried beancurd intervention <sup>[20]</sup>. However, these studies did not conduct extensive screening of urinary metabolites in response to soy intake. Additionally, the specificity and robustness of these biomarkers may be influenced by metabolizing enzymes <sup>[21]</sup>, gut microbiota <sup>[22]</sup>, and variations in dietary combinations, cooking or processing methods <sup>[23]</sup>. Despite the prevalence of a diet combining cereal-based foods and soy products in China, few studies have investigated potential biomarkers for this combined intake.

To our knowledge, this study is the first dietary intervention trial investigating urinary biomarkers for the combined intake of whole-grain cereals (corn, millet, oats) and soybeans. Utilizing untargeted metabolomics approaches, we analyzed metabolite concentrations in 24-hour urine samples collected from participants following interventions with these foods. The objective of this study was to screen for differential metabolites across intervention groups to identify biomarkers for soybeans, corn, millet and oats intake. This approach aims to improve the accuracy of dietary exposure assessment and facilitate better estimation of diet-disease associations.

## 2. Methods

### 2.1 Study design

In this randomized cross-over trial, participants underwent four different breakfasts interventions during 8 consecutive weeks. Each intervention lasted one day, preceded by a 5-day run-in period and followed by an 8-day washout period. Eligible participants were randomly assigned to one of four breakfast groups at each intervention, consuming special breakfast containing varying amounts of millet, corn, oats and soy products. Urine samples were collected at 1, 2, 3, 4, 6, 8, 12, and 24 hours after the meal. These samples were subsequently pooled for untargeted metabolomic analysis (Flowchart shown in Fig. 1). One week before the formal trial, the participants' physical and hematological parameters were measured and their daily dietary intake was investigated using a food frequency questionnaire (FFQ). Participants' diet and health status were also regularly monitored and recorded throughout the study. The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Capital Medical University (Approved number: Z2020SY044). All participants signed an informed consent form for the intervention.



**Fig 1.** Flow chart of intervention and urine sample collection

## 2.2 Study participants

To our knowledge, controlled human intervention studies incorporating metabolomic analyses remain very limited. Garcia-Perez et al. <sup>[12]</sup>, using changes in urinary proline betaine (a biomarker of citrus intake) as an example of a measurable dietary response, reported that a mean difference of 50  $\mu\text{mol/L}$  with a standard deviation (SD) of 40  $\mu\text{mol/L}$  could be detected with 12 participants (power = 0.95, alpha = 0.05). Given the complexity of multivariate models in metabolomics analysis compared to univariate models, we informed data from our prior preliminary study and followed the methodology outlined by van Iterson et al. <sup>[24]</sup>. Finally, it was calculated that 18 subjects were necessary to effectively identify differential urinary metabolites of different dietary pattern under power of 0.9 and a false discovery rate (FDR) of 0.1. To account for an anticipated 20% attrition rate, 4 additional subjects were recruited. To ensure a balanced sex ratio of 1:1 across the four intervention cohorts, a total of 24 participants were enrolled.

Participants were recruited through advertisements posted at Capital Medical University and its affiliated hospitals. Interested individuals underwent an initial screening to assess their eligibility based on the study's inclusion and exclusion criteria. During this process, members of the research team provided detailed explanations of the study protocol, procedures, potential risks and benefits, as well as participant management guidelines. A total of 24 participants aged between 18 and 20 years with a body mass index (BMI) ranging from 18.5 to 29.9 kg/m<sup>2</sup> were included in the study. Their weight change was limited to less than 3 kg during the preceding three months. Additionally, all the participants had no history of chronic disease (diabetes, cancer, cardiovascular and cerebrovascular diseases, kidney disease, liver disease, gastrointestinal diseases, pancreatitis) and medication. Other inclusion criteria included not smoking and not taking nutritional supplements in the last three months; The woman was not pregnant or breastfeeding; non-vegetarian or on a special diet.

### 2.3 Randomised control scheme and blinding

A random sequence was generated by Researcher A using the PROCPLAN procedure of the SAS software, specifying a seed number of 9683473, stratifying the randomisation by age and sex, and generating the random sequence. Researcher A was responsible for grouping the random sequences and was not involved in subject recruitment. The grouping information was placed in an opaque envelope by Researcher A, which was opened by Researchers B and C on the day of the dietary intervention and the next step in the trial.

### 2.4 Dietary intervention protocol

This study designed four types of test breakfasts (labeled A, B, C, and D), each containing different doses of millet, corn, oats and soy products (tofu, soymilk, dried tofu) as detailed in Table 1. Participants were randomly allocated to one of the breakfast groups ( $n = 6$  per group). Following four cross-over interventions, each participant had consumed A-D four different breakfasts. Each intervention was preceded by a 5-day run-in period and followed by a 8-day washout period. On the trial day, participants were asked to arrive at the test site before 8:00 a.m. They were required to consume all provided intervention breakfast on-site within 15 minutes. All foods were weighed immediately before serving, and any uneaten food was subsequently weighed to determine actual intake. Urine samples were collected at 1, 2, 3, 4, 6, 8, 12, and 24 hours after the meal (Fig. 1).

**Table 1.** Intakes of whole grains and soy products in the intervention groups <sup>a</sup>

|                 | Breakfast A | Breakfast B | Breakfast C | Breakfast D |
|-----------------|-------------|-------------|-------------|-------------|
| Millet, g       | 0           | 75          | 50          | 25          |
| Corn, g         | 75          | 50          | 0           | 100         |
| Oats, g         | 75          | 0           | 25          | 50          |
| Soy products, g | 125         | 215         | 325         | 0           |
| Tofu, g         | 50          | 75          | 100         | 0           |
| Dried tofu, g   | 25          | 40          | 75          | 0           |
| Soymilk, g      | 50          | 100         | 150         | 0           |
| Carbohydrate, g | 91.8        | 88.3        | 66.6        | 101.8       |

|                  |        |        |        |        |
|------------------|--------|--------|--------|--------|
| Fat, g           | 10.6   | 9.4    | 11.8   | 6.7    |
| Protein, g       | 28.8   | 28.5   | 35.3   | 18.4   |
| Dietary fiber, g | 9.7    | 6.1    | 4.9    | 9.4    |
| Energy, kj       | 2255.6 | 2205.0 | 2070.2 | 2106.2 |

Note: <sup>a</sup> All food weights represent dry weights of edible portions (except for soymilk). Foods were prepared by boiling with salt.

To minimize baseline dietary confounding, investigators provided all meals (breakfast, lunch, dinner) on the day before each intervention, as well as lunch and dinner on the trial day. These provided meals excluded any whole-grain cereals and soy products. Participants were instructed to refrain from consuming any additional foods after dinner on these days. Given that water intake significantly affects urine composition, participants were permitted to drink 100 mL of purified water with meals during each dietary intervention. Post-meal water intake was restricted to no more than 100 mL per hour.

A 5-day run-in period was implemented to eliminate residual metabolites from participants' baseline diets, followed by an 8-day washout period to minimize residual effects between consecutive interventions. This design was implemented because most food-derived metabolites in urine typically peak at 2-8 hours after ingestion and return to baseline levels within 24-48 hours [14,25]. During the run-in and washout period period, participants were prohibited from consuming: 1) whole grains (millet, corn, oats, brown rice, whole wheat, barley, rye, oats, buckwheat, etc.); 2) legumes (mung beans, red beans, fava beans, peas, kidney beans, lentils, cowpeas, etc.); 3) soybeans and their products (dried bean curd, veggie chicken, tofu, soymilk, soy milk, soya bean sprouts, etc.); 4) alcohol. They were also restricted from engaging in high-intensity physical activity. Additionally, we developed a dietary recording application to monitor participants' daily food and water intake. This tool enabled real-time tracking of compliance with the prescribed run-in and washout dietary regimens throughout the study period. All aspects of dietary intake and intervention delivery were supervised by a professional clinical dietitian.

Criteria for exclusion of subjects during the intervention process: 1)those who changed their daily diet for personal reasons; 2)those who had adverse reactions after the dietary intervention; 3)those who could not continue to participate in the study due to the occurrence of other diseases; 4)those who had poor adherence and could not strictly follow the intervention plan; 5)those who did not return on time for the visit or missed the visit; and those who were not willing to continue to cooperate with the intervention.

## 2.5 Untargeted metabolomic analysis

We utilized 24-hour pooled urine samples for untargeted metabolomic analysis. The 24-hour urine samples were prepared by pooling urine samples based on the excretion volume ratio at designated time points, thoroughly mixed and dispensed into sterile centrifuge tubes for storage at -80 °C. For each 100 µL of urine sample, 300 µL of extract solution (methanol, containing isotopically-labelled internal standard mixture) was added. The reagents and chemicals were detailed in Supporting Table S1. The internal standard mixture was obtained from BioTree (Shanghai, China) and detailed in Supporting Table S2. The mixture were vortexed for 30 s, sonicated in an ice-water bath for 10 min and incubated at -40 °C for 1 h to precipitate proteins. The sample was then centrifuged at 12000 r/min (RCF= 13800 (×g), R= 8.6 cm) for 15 min at 4 °C. The resulting

supernatant was transferred to a fresh glass vial for analysis. The quality control (QC) sample was prepared by mixing an equal aliquot of the supernatants from all of the samples.

LC-MS/MS analysis was performed using a UHPLC system (Vanquish, Thermo Fisher Scientific, Waltham, MA, USA) with a UPLC HSS T3 column (2.1 mm × 100 mm, 1.8 μm) coupled to an Orbitrap Exploris 120 mass spectrometer (Orbitrap MS, Thermo, Waltham, MA, USA). The mobile phase consisted of 5 mmol/L ammonium acetate and 5 mmol/L acetic acid in water (A) and acetonitrile (B). The auto-sampler temperature was 4 °C, and the injection volume was 2 μL. The Orbitrap Exploris 120 mass spectrometer was operated in an information-dependent acquisition (IDA) mode. Then, the acquisition software (Xcalibur 4.4, Thermo, Waltham, MA, USA) was used to evaluate the full-scan survey MS data based on preselected criteria. The ESI source conditions were set as follows: sheath gas flow rate as 50 Arb, aux gas flow rate as 15 Arb, capillary temperature as 320 °C, full MS resolution as 60000, MS/MS resolution as 15000, MS/MS intensity as 200000, MS scan rate as 3 Hz, collision energy as 10/30/60 in NCE mode, spray voltage as 3.8 kV (positive) or -3.4 kV (negative), respectively. 4 ions per cycle were fragmented.

## 2.6 Quality control

System suitability tests were routinely performed. QC samples were interspersed within the analytical sequence (one QC sample per 12 experimental urine samples) to provide data from which the reproducibility could be assessed. QC samples clustered closely together in the principal component analysis (PCA) score plots confirming a stable system (Supporting Fig. S1A, B). The high correlation of the QC samples indicated the high quality of the data from this experiment (Fig. S1C, D). Additionally, the observed median RSD for internal standards in QC samples was ≤ 10%, reflecting consistent internal standard performance.

## 2.7 Data preprocessing and metabolites annotation

The raw data were converted to the mzXML format by ProteoWizard and processed by R package XCMS (version 3.6.2, ppm:10, S/N:3, peak width:5-20). The process includes peak detection, extraction, alignment, and integration. The peaks with RSD% (calculating in the QC samples) > 30% were removed. Also, the peaks with missing values in more than 50% of samples were filtered out. For the left missing values, the data were filled up with half of the minimum value.

Metabolite annotation was based on four key parameters: (1) retention time (rt) (2) accurate mass measurement (mass-to-charge ratio,  $m/z$ ) (3) MS/MS fragmentation (4) spectral matching against an in-house MS/MS spectral database (BiotreeDB, Shanghai), which contains over 2,000 authenticated reference standards. MS/MS similarity score ( $MS^2$  score) was calculated via forward dot-product algorithm [26], reflecting the similarity between the spectra of the unknown compounds and MS/MS spectral database. A higher  $MS^2$  score (ranging from 0-1) indicates greater confidence in identification, with smaller errors in spectral matching. We employed a threshold of > 0.8 to ensure high-confidence metabolite annotations and maintain FDR at an acceptable level [27]. In total, 380 high-confidence metabolites were identified in positive ionization mode from urine samples and 195 metabolites in negative ionization mode. Subsequently,

according to the latest identification confidence levels proposed in Nature Methods <sup>[28]</sup>, metabolites were classified into four confidence categories (A-D): (A) standard or NMR; (B(i)) confident match based on MS/MS and (B(ii)) confident match using in-silico MS/MS approaches and (B(iii)) partial match based on MS/MS and (C(i)) confident match based on MSn and (C(ii)) confident match using in-silico MSn approaches and (C(iii)) partial match based on MSn; (D) MS only. All metabolites identified in this study were based on reliable MS/MS matching. In positive ionization mode, 54 metabolites were classified as B(i) and 326 metabolites as B(ii). In negative ionization mode, 99 metabolites were classified as B(i) and 96 metabolites as B(ii).

## 2.8 Statistical analysis

Data was expressed as mean and SD unless otherwise stated. Preprocessed peak area data was normalized by volume of 24-hour pooled urine and log<sub>10</sub>-transformed. Univariate analysis (paired t-tests) was used to compare differences between the two groups,  $P < 0.05$  was considered to indicate statistical significance. p-value was corrected for multiple testing by the Benjamini–Hochberg FDR. Fold-change (FC) analysis was employed to evaluate the absolute differences in metabolite mean values between the two experimental groups. Significantly differential metabolites were identified by combining FC analysis and t-tests, applying dual thresholds of  $|\log_2\text{FC}| > 1$  (equivalent to  $\text{FC} > 2$  or  $< 0.5$ ) and  $\text{FDR} < 0.05$ . To carry out multivariate analysis, within-individual effect matrix was extracted as described by Westerhuis et al <sup>[29]</sup>, using R package mixOmics <sup>[30]</sup>. Random forest classification models were then built to classify soybean product group or corn group based on metabolite profile <sup>[31]</sup>. Three different consensus models (“min”, “mid” and “max”) were build using area under the receiver operation characteristics curve (AUROC) as metrics. In general, nRep, nOuter and varRatio were set to be 100, 20, 0.75, respectively. The ‘min’ and ‘max’ models represented the model with minimal-optimal and all-relevant predictors, respectively. And ‘mid’ corresponded to the geometric mean of the max and min number of variables. The number of variables of ‘min’ and ‘max’ model was determined based on averaged validation performance per repetition and overall, respectively. In addition, the permutation tests were conducted by repeated random sampling ( $n = 100$ ) of the original intervention group variables and thereafter modelling original metabolites matrix on permuted intervention group responses, thus obtaining a population of fitness metrics corresponding to a null hypothesis distribution. It was then assumed that the population of fitness metrics from random permutations had a Gaussian t-distribution and permutation p-values were calculated as the cumulative 1-tailed probability of achieving the actual model fitness in the t-distribution <sup>[32]</sup>. Both results of univariate analysis and multivariate analysis were considered to decide differential metabolites. Linear regression analysis was performed with GraphPad Prism 9.0 between the amount of soy products, millet, corn consumed and the average concentration for each differential metabolites, yielding a regression equation and  $R^2$  values.

## 3. Results

### 3.1 Baseline Characteristics

Of the 24 subjects initially enrolled in the study, 21 completed all procedures. Three participants withdrew prior to completion: two did not finish the full intervention, and one was excluded due to insufficient urine output during the third and fourth collection periods. Baseline characteristics of the 21 completers are presented in Table 2. The cohort was predominantly of normal weight (95.2%), with one participant classified as overweight (BMI = 29.8 kg/m<sup>2</sup>).

**Table 2.** Baseline characteristics of participants <sup>a</sup>

|                        | Date (n=21)                           |
|------------------------|---------------------------------------|
| Sex                    |                                       |
| Male                   | 10 (48%)                              |
| Female                 | 11 (52%)                              |
| Age, year              | 19.0 (0.8; 18 - 20)                   |
| Weight, kg             | 58.6 (10.5; 46.1 – 92.0)              |
| BMI, kg/m <sup>2</sup> | 21.1 (2.8; 18.5 – 29.8)               |
| Waist-hip ratio        | 0.82 (0.04; 0.78 – 0.95)              |
| Body fat percentage, % | 19.2 (6.4; 10.4 – 28.6)               |
| Glucose, mmol/L        | 4.46 (0.33; 3.84 - 5.06)              |
| HbA1c, %               | 5.3 (0.2; 4.9 – 5.7)                  |
| Triglycerides, mmol/L  | 0.84 (0.53; 0.35 – 2.99)              |
| Cholesterol, mmol/L    |                                       |
| Total                  | 4.29 (0.67; 3.50 - 5.92)              |
| LDL                    | 2.41 (0.57; 1.48 – 3.43)              |
| HDL                    | 1.69 (0.36; 1.04 – 2.61)              |
| Whole grains, g/d      | 13.64 (13.80; 0 – 50.00 )             |
| Soy products, g/d      | 61.26 (69.96; 0 – 255.00)             |
| Energy, kJ/d           | 7267.27 (2817.09; 3258.88 – 15698.95) |
| Carbohydrate, g/d      | 165.16 (74.74; 63.57 – 404.93)        |
| Fat, g/d               | 92.52 (30.28; 51.09 – 173.87)         |
| Protein, g/d           | 64.47 (32.19; 17.51 – 153.33)         |
| Dietary fiber, g/d     | 5.58 (3.75; 1.89 – 17.80)             |

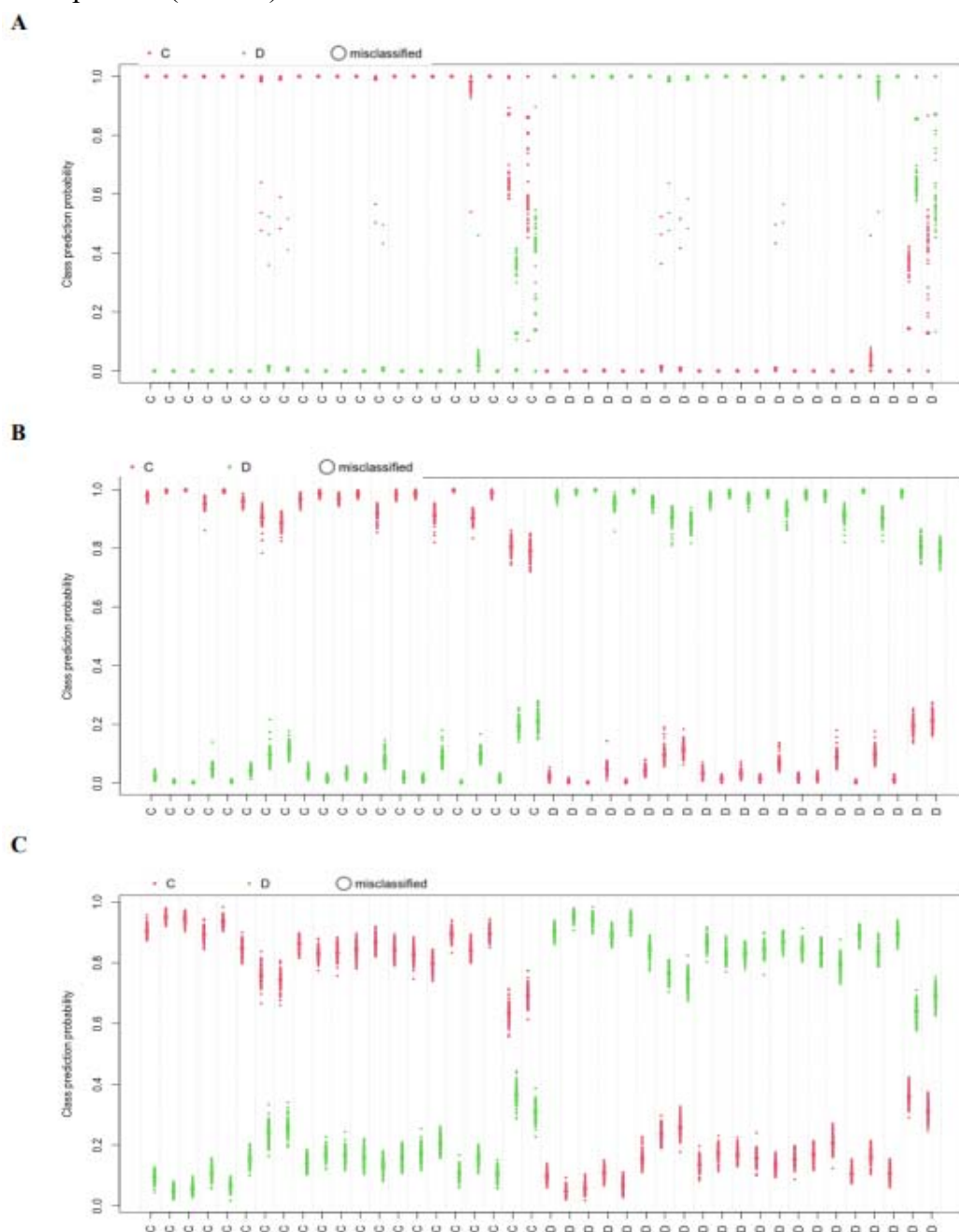
Note: <sup>a</sup> Data are n (%) or mean (SD; range). HbA1c: Glycated hemoglobin A1c; LDL, low-density lipoprotein; HDL, high-density lipoprotein

The dietary interventions were designed to contrast specific food components. Breakfasts A and B were formulated to differ primarily in their oat (75 g vs. 0 g) and millet (0 g vs. 75 g) content. Similarly, Breakfast C and D were distinguished by their corn (0 g vs. 100 g) and soy products (325 g vs. 0 g) content. Consequently, urinary metabolites from Breakfast A and B were compared to identify potential biomarkers for oat and millet intake. Likewise, metabolites from Breakfast C and D were compared to identify biomarkers for corn and soy products consumption.

### 3.2 Identifying Differential Metabolites by Univariate and Multivariate Analyses

Significantly differential metabolites were identified using both univariate and multivariate analyses. In the comparison between group A and B (millet vs. oats), univariate analysis (t-test with FDR < 0.05 and FC > 2 or < 0.5) did not yield any significantly differential metabolites (Supporting Table S3). FC of riboflavin, flazine and genistein 5-O-glucuronide were 0.8, 0.7 and 0.6, respectively. However, multivariate random

forest analysis identified the "max" model as optimal for separating these groups (Supporting Fig. S2), with flazine emerging as a feature of notable importance (average rank = 12.2). In contrast, the comparison between group C and D (soy products vs. corn) yielded 13 significantly differential metabolites that met the statistical thresholds ( $\text{FDR} < 0.05$ ,  $\text{FC} > 2$  or  $< 0.5$ ; Supporting Table S4). We additionally identified 5 metabolites meeting the criterion of  $\text{FDR} < 0.05$  with FC values ranging from 0.6 to 0.8. For multivariate analysis of these groups, the "mid" random forest model was selected as optimal based on class prediction probabilities (Fig. 2), despite multiple models achieving a perfect AUROC of 1.0 (Supporting Table S5). Finally, we identified 3 candidate differential metabolites for the group A vs. B comparison (Supporting Table S3) and 14 for the group C vs. D comparison (Table 3).



**Fig. 2.** Class prediction probability plot of three consensus model using Random Forest in breakfast C and D.

**Table 3.** Differentially expressed urinary metabolites in breakfast C and D identified using both univariate and multivariate analyses <sup>a</sup>

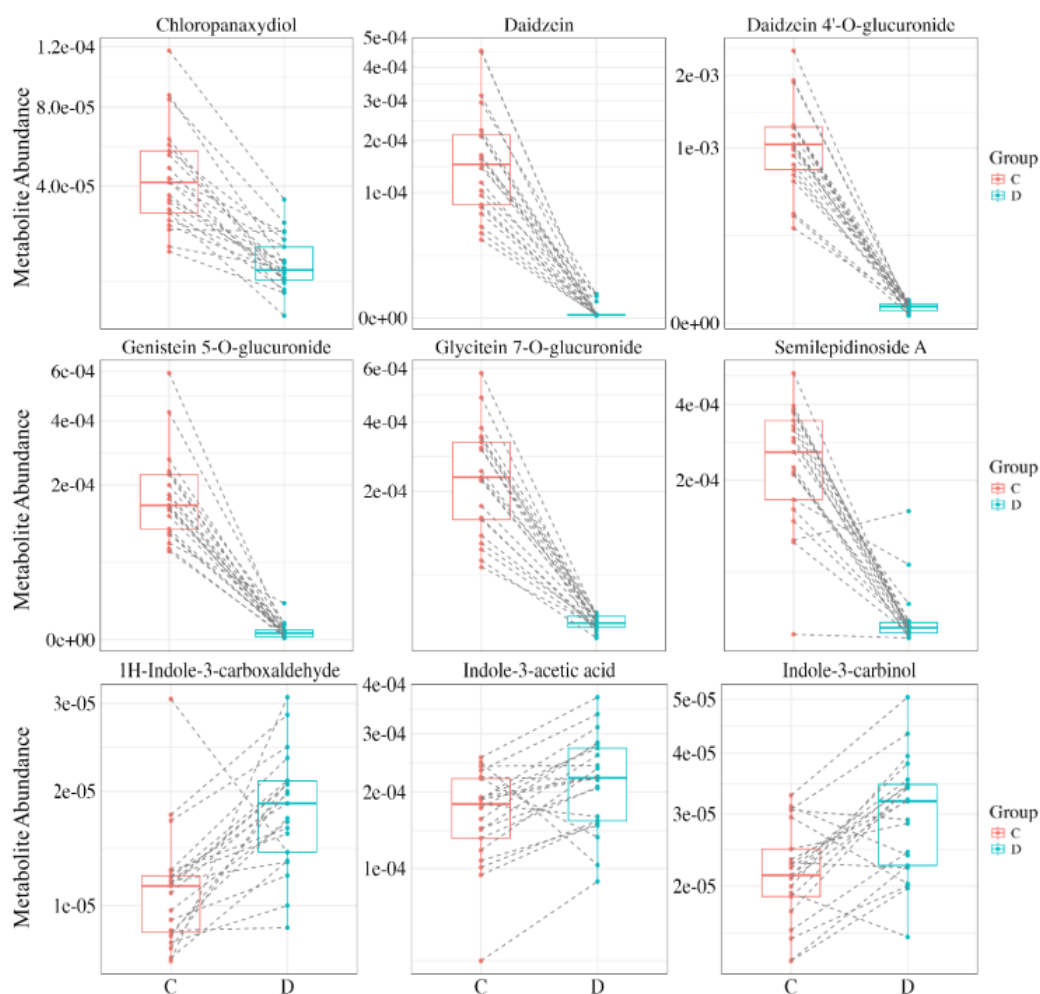
| Metabolites                | RT<br>(min) | m/z      | MS <sup>2</sup><br>score | Identification<br>level <sup>b</sup> | Breakfast C <sup>c</sup> |         | Breakfast D <sup>c</sup> |         | Fold<br>change | P       | FDR     | Mean<br>rank | Order |
|----------------------------|-------------|----------|--------------------------|--------------------------------------|--------------------------|---------|--------------------------|---------|----------------|---------|---------|--------------|-------|
|                            |             |          |                          |                                      | Mean                     | SD      | Mean                     | SD      |                |         |         |              |       |
| Daidzein                   | 5.85        | 255.0644 | 0.92                     | B(i)                                 | 1.7E-04                  | 1.2E-04 | 6.1E-07                  | 1.2E-06 | 284.0          | 1.6E-11 | 7.6E-10 | 3.4          | 5     |
| Genistein 5-O-glucuronide  | 4.69        | 447.0914 | 0.96                     | B(ii)                                | 1.9E-04                  | 1.3E-04 | 1.1E-06                  | 2.4E-06 | 167.9          | 6.3E-12 | 4.0E-10 | 3.4          | 4     |
| Daidzein 4'-O-glucuronide  | 4.13        | 431.0963 | 0.95                     | B(ii)                                | 1.1E-03                  | 5.2E-04 | 8.7E-06                  | 4.7E-06 | 123.6          | 2.8E-16 | 1.1E-13 | 3.3          | 1     |
| Glycitein 7-O-glucuronide  | 4.21        | 461.1073 | 0.96                     | B(ii)                                | 2.5E-04                  | 1.4E-04 | 9.4E-06                  | 3.5E-06 | 26.4           | 7.5E-14 | 5.7E-12 | 3.4          | 2     |
| Semilepidinoside A         | 3.21        | 337.1463 | 1.00                     | B(ii)                                | 2.6E-04                  | 1.3E-04 | 1.6E-05                  | 3.0E-05 | 15.8           | 5.9E-14 | 5.6E-12 | 7.7          | 7     |
| Chloropanaxydiol           | 4.09        | 313.1619 | 0.85                     | B(ii)                                | 4.6E-05                  | 2.5E-05 | 1.5E-05                  | 7.3E-06 | 3.2            | 1.0E-11 | 5.4E-10 | 3.4          | 3     |
| Nigakihemiacetal B         | 4.86        | 391.2195 | 0.98                     | B(ii)                                | 3.4E-04                  | 1.9E-04 | 1.2E-04                  | 7.2E-05 | 2.9            | 4.0E-11 | 1.7E-09 | 11.0         | 13    |
| Riboflavin                 | 3.66        | 377.1447 | 0.99                     | B(ii)                                | 4.9E-05                  | 1.3E-05 | 1.8E-05                  | 1.1E-05 | 2.7            | 9.1E-15 | 1.7E-12 | 9.2          | 8     |
| Biotin sulfone             | 3.12        | 277.0786 | 0.84                     | B(ii)                                | 7.9E-05                  | 2.5E-05 | 3.0E-05                  | 2.1E-05 | 2.6            | 2.8E-14 | 3.5E-12 | 5.4          | 6     |
| Ovalicin                   | 5.48        | 297.1667 | 0.98                     | B(ii)                                | 1.8E-04                  | 9.0E-05 | 7.5E-05                  | 3.8E-05 | 2.4            | 8.7E-10 | 3.0E-08 | 11.0         | 12    |
| 10-Methoxyyangonin         | 4.99        | 289.1018 | 0.97                     | B(ii)                                | 1.3E-04                  | 5.9E-05 | 6.1E-05                  | 3.9E-05 | 2.1            | 7.8E-09 | 2.3E-07 | 10.9         | 11    |
| Indole-3-acetic acid       | 4.24        | 176.0700 | 1.00                     | B(i)                                 | 1.7E-04                  | 5.9E-05 | 2.2E-04                  | 7.5E-05 | 0.8            | 3.7E-05 | 5.4E-04 | 129.1        | 25    |
| Indole-3-carbinol          | 4.23        | 130.0647 | 1.00                     | B(ii)                                | 2.2E-05                  | 6.3E-06 | 3.0E-05                  | 8.9E-06 | 0.7            | 2.0E-07 | 5.2E-06 | 29.9         | 18    |
| 1H-Indole-3-carboxaldehyde | 4.21        | 146.0596 | 0.96                     | B(ii)                                | 1.2E-05                  | 5.3E-06 | 1.8E-05                  | 5.7E-06 | 0.6            | 4.6E-07 | 1.1E-05 | 10.9         | 10    |

Note: <sup>a</sup> RT, retention time; m/z, exact mass-to-charge ratio measured; MS<sup>2</sup> score: MS/MS similarity score; Mean: mean relative quantification values of metabolites in an experimental group; SD, standard deviation; FDR, false discovery rate; Mean rank, the average importance ranking of metabolites in the random forest model; Order, the importance ordering of metabolites in the model. The preprocessed peak area data obtained from positive ionization mode was normalized by 24-hour pooled urine volume. P-value was corrected for multiple testing by the Benjamini–Hochberg FDR.

<sup>b</sup> Identification level (A; B; C; D): (A) standard or NMR; (B(i)) confident match based on MS/MS and (B(ii)) confident match using in-silico MS/MS approaches and (B(iii)) partial match based on MS/MS and (C(i)) confident match based on MSn and (C(ii)) confident match using in-silico MSn approaches and (C(iii)) partial match based on MSn; (D) MS only

<sup>c</sup> Breakfast C included 325g soy products and 0g corn, breakfast D included 0g soy products and 100g corn.

The individual trajectories of key metabolites for the group C and D comparison are illustrated in Fig. 3. The relative abundances of four isoflavones (daidzein, daidzein 4'-O-glucuronide, genistein 5-O-glucuronide, and glycitein 7-O-glucuronide) were nearly zero in group D (0 g soy intake). Correspondingly, all participants showed a sharp decrease in these metabolites from group C (325 g soy intake) to group D, confirming a direct response to soy removal. The decline for other potential soy-related metabolites, such as chloropanaxydiol and semilepidinoside A, was less consistent across participants. Conversely, three indole compounds (indole-3-acetic acid (IAA), indole-3-carbinol (I3C), and 1H-indole-3-carboxaldehyde (IAld)) increased from group C (0 g corn intake) to group D (100 g corn intake). The majority of participants showed a clear increasing trend, although a few individuals exhibited an opposite response, indicating inter-individual variability.

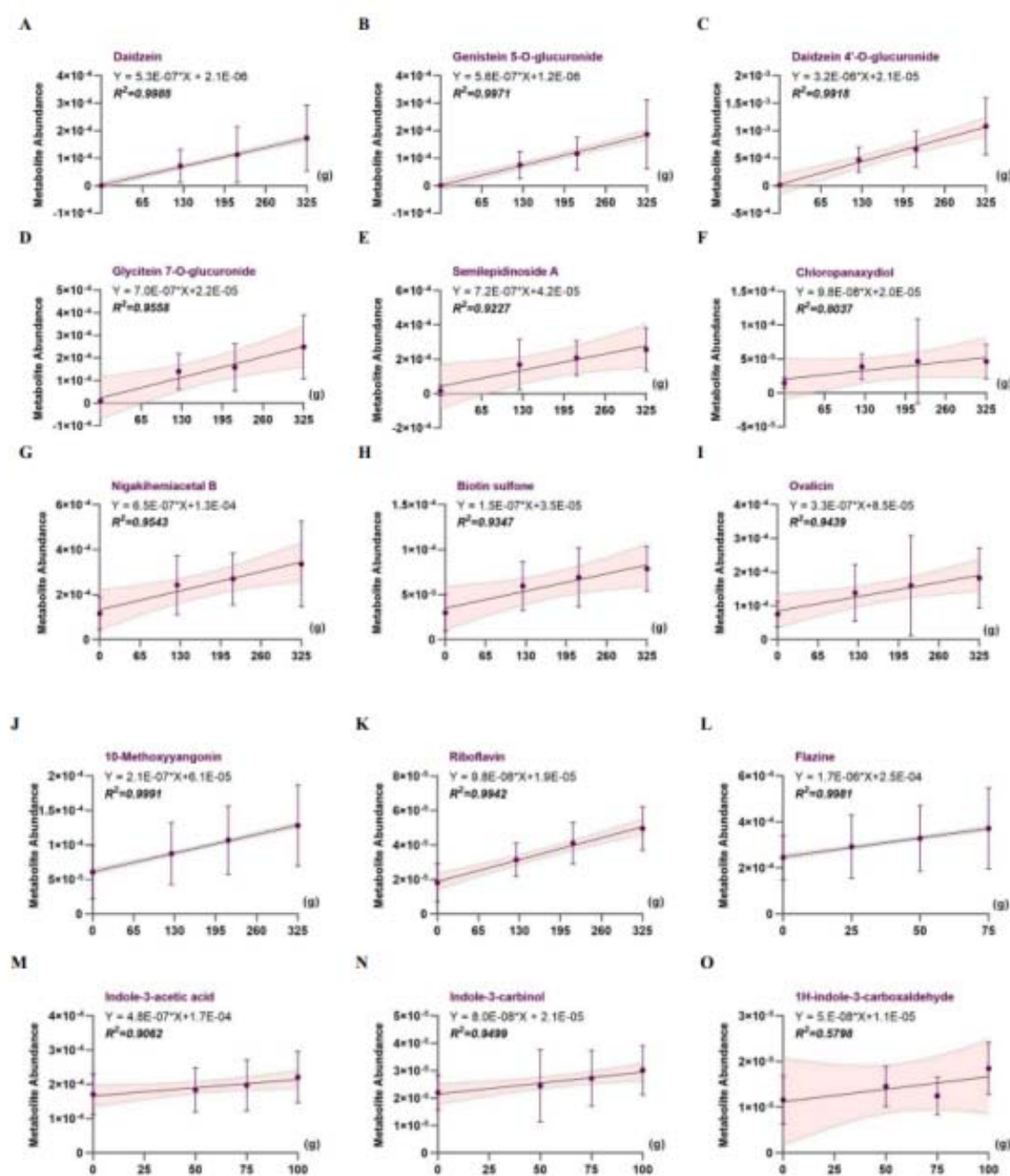


**Fig. 3.** Individual trajectories of differential metabolites in breakfast C and D.

### 3.3 Dose-Response Relationships of Metabolites with Food Intake

Linear regression was used to model the dose-response relationship between food intake and the relative abundance of differential metabolites (Fig. 4). Eleven metabolites showed a dose-dependent increase with soy products intake (0, 125, 215, and 325 g). Four isoflavones, in particular, demonstrated exceptionally strong linear correlations ( $R^2 = 0.96\text{--}0.99$ ) and were nearly undetectable at zero soy intake (Fig. 4A-D). These results,

combined with their consistent individual trajectories (Fig. 3), identify them as strong candidate biomarkers for soy consumption. Riboflavin also showed a strong correlation with soy intake ( $R^2=0.99$ ; Fig. 4K). Flazine relative abundance was strongly correlated with millet intake ( $R^2 = 0.99$ ; Fig. 4L). However, inconsistent trajectories in some individuals (Supporting Fig. S3) suggested its utility as a millet biomarker required further validation. Finally, IAA, I3C, and IAld levels were positively correlated with corn intake ( $R^2=0.91, 0.95$ , and  $0.58$ , respectively; Fig. 4M-O).



**Fig. 4.** Dose-response of soy products, millet and corn.

Results from the negative ionization mode are provided in the Supplementary Files (Supporting Fig. S4-7). Notably, daidzein showed a strong linear correlation with soy product intake ( $R^2=0.99$ ; Supporting Fig. S8). The pairwise group comparisons confirmed significant changes between diets high and low in soy (Supporting Table S6, S7), further supporting its role as a robust soy intake biomarker.

## 4. Discussion

Isoflavones including genistein, glycitein and daidzein have been extensively documented as biomarker of soybean and soybean product intakes [8,33]. Our study identified three isoflavone O-glucuronides (genistein 5-O-glucuronide, daidzein 4'-O-glucuronide and glycitein 7-O-glucuronide) as well as daidzein (with fold change over 25) as potential biomarkers of soybean intakes. Particularly, daidzein showed a strong correlations with soy intake in both positive and negative ionization modes. Previously, isoflavone glucuronides have been reported to be the dominant isoflavone metabolites in human urine and plasma after consumption of soymilk powder, which made up 70-80% and 50-60% of isoflavones in urine and plasma samples, respectively [34]. Interestingly, in another in vivo study based on rat model, urinary isoflavonoid O-glucuronides were identified as black soybean metabolites [35], where glucuronide conjugates of daidzein and genistein (such as daidzein-7-O-glucuronide, daidzein-4'- $\beta$ -D-glucuronide and genistein-5-O-glucuronide etc.) showed high intensity in rat urine samples [35]. The composition and profile of soy isoflavone isomers are significantly affected by food processing procedure [36]. Our study reinforce and expand the evidence to a real-world dietary setting combining soy with whole-grain cereals. Furthermore, our study included three different soybean products (tofu, dried tofu and soymilk) which make up the major soy foods intakes in the Chinese population, providing a more extensive assessment of dietary exposure.

Riboflavin (vitamin B2) is a water-soluble vitamin of the B complex group. In our study, riboflavin relative abundance demonstrated a remarkably strong correlation ( $R^2 = 0.99$ ) with soy products intake. We observed 9.2-fold increase in riboflavin relative abundance following consumption of 325 g soy products compared to zero intake. Previous studies have shown high levels of riboflavin in fermented soy products, where microbial activity during fermentation enhances riboflavin synthesis. For instance, *Lactobacillus plantarum* CRL 2130 has been shown to produce riboflavin during soymilk fermentation, with the fermented soymilk effectively alleviating riboflavin deficiency symptoms in murine models [37]. But riboflavin can be found in a wide variety of foods and natural sources (especially milk, organ meats, dark green leafy vegetables) [38]. Therefore, the utility of riboflavin as a specific biomarker for soy intake is limited.

Urine semilepidinoside A relative abundance increased over 15 fold following consumption of 325 g soybean products in our study. Semilepidinoside A belongs to the class of organic compounds known as phenolic glycosides, which has been identified in vegetable including brassicas and garden cresses (*Lepidium sativum*) [39] but not in soybean. Chloropanaxydiol, a chlorohydrin compound first isolated from *Panax ginseng* [40], has shown potential as a diagnostic biomarker for sepsis in a case-control study [41], but its specificity for soy products appears limited. The biotin derivative biotin sulfone, while detectable, originates from multiple dietary sources including egg yolk, liver, cereals and vegetables [42], compromising its specificity as a soy marker. Fungal-derived ovalicin [43] and the rarely detected nigakihemiacetal B and 10-methoxyyangonin all show insufficient evidence for association with soy consumption in current literature. These findings collectively suggest that while certain metabolites demonstrate quantitative responses to soy intake, their non-

specific distribution across various biological sources necessitates caution in interpreting them as definitive soy biomarkers without further validation studies.

Research on biomarkers of dietary corn or maize intake remains limited. In the present study, we identified three indole derivatives (IAA, I3C and IAld) as potential biomarkers associated with corn intake. Nevertheless, these indole derivatives are not unique to corn and can originate from other dietary sources<sup>[44,45]</sup> or endogenous metabolism<sup>[46]</sup>. For instance, indole-2-carboxylic acid, 3-indolecarboxylic acid glucuronide and indole-3-lactic acid have been documented to be biomarker of rice bran, whole grain bread and fermented food intakes, respectively<sup>[47,48]</sup>. Furthermore, tryptophan (Trp) serves as the primary endogenous precursor for indole derivatives, which can be transformed by various human gut microbiota (e.g., *Lactobacillus* and *Bifidobacterium*)<sup>[49]</sup> into structurally similar yet functionally diverse indole compounds, including the three indole derivatives identified in our study<sup>[50-52]</sup>. To minimize the potential influence of confounding factors, strict dietary controls were implemented during the experimental period by providing participants with standardized meals that excluded cruciferous vegetables and other known indole precursors sources. Our study showed a significant increase in levels of indole derivatives (IAA, I3C and IAld) from breakfast C (zero corn intake) to breakfast D (100 g corn intake), with IAA and I3C exhibiting strong dose-response relationships with corn consumption. This pattern provides compelling evidence supporting their potential as corn-specific biomarkers. Previous studies have also found high quantities of IAA conjugates in maize endosperm, making corn a suitable model system for analyzing IAA biosynthesis<sup>[53]</sup>. Moreover, among eight edible vegetable oils, IAld and 1H-indole-3-ethyl acetate were detected exclusively in corn oil<sup>[54]</sup>. These evidences further support that indole derivatives have a particular link with corn. We also observed interindividual variability in urinary indole metabolite trajectories, with a few participants exhibiting elevated metabolite levels even under conditions of low corn intake. We hypothesize this variation may be attributed to differences in gut microbial metabolic activity<sup>[55]</sup>. To further establish the corn specificity of these markers, targeted metabolomics and validation in larger cohorts with controlled dietary backgrounds are warranted. Future studies should also investigate the stability of these candidate biomarkers across diverse populations and dietary patterns to confirm their specificity and applicability in assessing corn intake.

Among all candidate metabolites screened in our study, flazine emerged as the only compound demonstrating a consistent trend with millet intake, showing a modest but significant 1.4-fold increase from breakfast A (zero millet intake) to breakfast B (75 g millet intake). While flazine was originally isolated from the roots of *Anemone altaica*<sup>[56]</sup>, it has also been isolated from methanol extract of kanjang (Korean soy sauce), a byproduct of the production of the Korean fermented soybean<sup>[57]</sup>. Additionally, a previous metabolomic study on millet from China, did not report flazine as a differential metabolite<sup>[58]</sup>. Further validation studies are needed to establish its specificity and dose-response relationship in controlled dietary interventions, particularly in populations with varying intakes of both millet and fermented soy products.

Oats contain two unique classes of compounds: avenanthramides (AVAs), avenacosides (AVEs), and their metabolites may serve as potential short- to medium-term biomarkers of oat intake<sup>[59-60]</sup>, as they are not

present in other commonly consumed foods <sup>[61-62]</sup>. Given that no specific oat biomarkers were identified in the current study, we considered that the intervention dose might be a contributing factor. Schär et al. <sup>[25]</sup> examined the excretion of phenolic acids and AVAs in urine samples from subjects after consumption of 60 g of oat bran. They observed increased levels of a total of 30 phenolic compounds within 8 hours after intake. Similarly, Wang et al. <sup>[60]</sup> reported a rapid increase in the urinary concentrations of AVE-A and B within 48 hours following ingestion of 120 g of oat bran. The oat intervention dose in our study (0-75 g) was comparable to those employed in these investigations. However, our study design reflected typical grain-soy combination patterns in the diet. Conventional intervention doses might be insufficient to highlight oat biomarkers in complex dietary patterns. We propose that higher intake levels or longer intervention periods may be necessary to detect specific biomarkers in such mixed-diet patterns.

The major limitation of the present study was that metabolite identification confidence primarily relies on in-house MS/MS spectral matching, without further validation using targeted metabolomic approach such as direct comparison with standards. Therefore, the identified metabolites remain exploratory, and their broader applicability requires further confirmation in large-scale and diverse studies involving populations of different age groups. Another limitation was that the metabolites and their potential precursors were not determined directly in food that being consumed, therefore the plausibility of those potential biomarkers (especially those of corn intake) should be further confirmed <sup>[63]</sup>. The lack of discriminatory metabolites for millet and oats may relate to our design reflecting grain-soy combinations patterns, where conventional doses might be insufficient to highlight their biomarkers rather than true absence of biomarkers. Higher intake levels or longer intervention periods should be tested in future studies.

## 5. Conclusion

This randomized cross-over trial identified daidzein and several isoflavone O-glucuronides (genistein 5-O-glucuronide, daidzein 4'-O-glucuronide and glycitein 7-O-glucuronide) as potential biomarker for soybean product intakes. Furthermore, it is essential to validate indole derivatives—such as indole-3-acetic acid, indole-3-carbinol, and indole-3-carboxaldehyde—across diverse populations and dietary patterns to confirm their specificity and applicability in the assessment of corn intake.

## Conflict of Interest

Authors declare that they have no conflict of interests.

## References

- [1] N.P.V. P, I.J. Joye, Dietary fibre from whole grains and their benefits on metabolic health, *Nutrients*. 12(10) (2020) 3045. <https://doi.org/10.3390/nu12103045>.
- [2] C.J. Seal, C.M. Courtin, K. Venema, et al., Health benefits of whole grain: effects on dietary carbohydrate quality, the gut microbiome, and consequences of processing, *Compr. Rev. Food. Sci. Food Saf.* 20(3) (2021) 2742-2768. <https://doi.org/10.1111/1541-4337.12728>.
- [3] G. Rizzo, L. Baroni, Soy, soy foods and their role in vegetarian diets, *Nutrients*. 10(1) (2018) 43. <https://doi.org/10.3390/nu10010043>.

- [4] G.N. Bookwalter, Soy protein utilization in food systems, *Adv. Exp. Med. Biol.* 105 (1978) 749-766. [https://doi.org/10.1007/978-1-4684-3366-1\\_36](https://doi.org/10.1007/978-1-4684-3366-1_36).
- [5] F.B. Hu, W.C. Willett, Current and future landscape of nutritional epidemiologic research, *JAMA*. 320(20) (2018) 2073-2074. <https://doi.org/10.1001/jama.2018.16166>.
- [6] M. Guasch-Ferre, S.N. Bhupathiraju, F.B. Hu, Use of metabolomics in improving assessment of dietary intake, *Clin. Chem.* 64(1) (2018) 82-98. <https://doi.org/10.1373/clinchem.2017.272344>.
- [7] R. Landberg, P. Karra, R. Hoobler, et al., Dietary biomarkers-an update on their validity and applicability in epidemiological studies, *Nutr. Rev.* 82(9) (2024) 1260-1280. <https://doi.org/10.1093/nutrit/nuad119>.
- [8] A.A. Franke, S.M. Hebshi, I. Pagano, et al., Urine accurately reflects circulating isoflavonoids and ascertains compliance during soy intervention, *Cancer. Epidemiol. Biomarkers. Prev.* 19(7) (2010) 1775-1783. <https://doi.org/10.1158/1055-9965.EPI-10-0116>.
- [9] S.X. Liang, H.H. Zhang, Y.F. Mo, et al., Urinary equol and equol-predicting microbial species are favorably associated with cardiometabolic risk markers in chinese adults, *J. Am. Heart Assoc.* 13(13) (2024) e034126. <https://doi.org/10.1161/JAHA.123.034126>.
- [10] H.M. Roager, J.K. Vogt, M. Kristensen, et al., Whole grain-rich diet reduces body weight and systemic low-grade inflammation without inducing major changes of the gut microbiome: a randomised cross-over trial, *Gut*. 68(1) (2019) 83-93. <https://doi.org/10.1136/gutjnl-2017-314786>.
- [11] Y.X. Yang, X.L. Wang, P.M. Leong, et al., New chinese dietary guidelines: healthy eating patterns and food-based dietary recommendations, *Asia Pac. J. Clin. Nutr.* 27(4) (2018) 908-913. <https://doi.org/10.6133/apjcn.072018.03>.
- [12] I. Garcia-Perez, J.M. Posma, R. Gibson, et al., Objective assessment of dietary patterns by use of metabolic phenotyping: a randomised, controlled, crossover trial, *Lancet Diabetes Endocrinol.* 5(3) (2017) 184-195. [https://doi.org/10.1016/S2213-8587\(16\)30419-3](https://doi.org/10.1016/S2213-8587(16)30419-3).
- [13] M. Jawhara, S.B. Sorensen, B.L. Heitmann, et al., Biomarkers of whole-grain and cereal-fiber intake in human studies: a systematic review of the available evidence and perspectives, *Nutrients*. 11(12) (2019) 2994. <https://doi.org/10.3390/nu11122994>.
- [14] P.S.C. Sri Harsha, R.A. Wahab, M. Garcia-Aloy, et al., Biomarkers of legume intake in human intervention and observational studies: a systematic review, *Genes Nutr.* 13 (2018) 25. <https://doi.org/10.1186/s12263-018-0614-6>.
- [15] Y.J. Yang, Y.L. Zhou, W.Y. Ye, et al., Integration of LC-MS-based untargeted and targeted metabolomics to uncover novel whole-grain wheat dietary intake biomarkers in the plasma of the chinese population, *Food Res. Int.* 202 (2025) 115740. <https://doi.org/10.1016/j.foodres.2025.115740>.
- [16] T.P. Sun, Y. Rong, X.L. Hu, et al., Plasma alkylresorcinol metabolite, a biomarker of whole-grain wheat and rye intake, and risk of type 2 diabetes and impaired glucose regulation in a chinese population, *Diabetes Care*. 41(3) (2018) 440-445. <https://doi.org/10.2337/dc17-1570>.
- [17] Z.S. Sun, Y.L. Shao, X.Y. Zhang, et al., Urinary alkylresorcinol metabolites, biomarkers of whole grain wheat and rye intake, are beneficially associated with liver fat and other fat measures, *Food Funct.* 15(16) (2024) 8488-8495. <https://doi.org/10.1039/d4fo01248h>.
- [18] A. Frank, H. Greve, F. Hubner, et al., Human intervention study: alkylresorcinol metabolites as potential biomarkers for grain intake and the occurrence of alkylresorcinols in commonly consumed foods in the german retail sector, *ACS Omega*. 9(22) (2024) 23555-23566. <https://doi.org/10.1021/acsomega.4c00733>.
- [19] Z.M. Liu, S.C. Ho, Y.M. Chen, et al., Whole soy, but not purified daidzein, had a favorable effect on improvement of cardiovascular risks: a 6-month randomized, double-blind, and placebo-controlled trial in equol-producing postmenopausal women, *Mol. Nutr. Food Res.* 58(4) (2014) 709-717. <https://doi.org/10.1002/mnfr.201300499>.
- [20] L. Li, M.J. Sun, J.P. Sun, et al., The effect of dried beancurd on bone mineral density in postmenopausal chinese women: a 2-year randomized controlled trial, *Calcif. Tissue Int.* 105(6) (2019) 573-581. <https://doi.org/10.1007/s00223-019-00604-2>.
- [21] H.J. Kim, M.S. Choi, S.U. Rehman, et al., Determination of urinary caffeine metabolites as biomarkers for drug metabolic enzyme activities, *Nutrients*. 11(8) (2019) 1947. <https://doi.org/10.3390/nu11081947>.

- [22] C. Favari, J.F. Rinaldi De Alvarenga, L. Sanchez Martinez, et al., Factors driving the inter-individual variability in the metabolism and bioavailability of (poly)phenolic metabolites: a systematic review of human studies, *Redox Biol.* 71 (2024) 103095. <https://doi.org/10.1016/j.redox.2024.103095>.
- [23] M. Rodriguez-Ayala, J.R. Banegas, R. Ortola, et al., Cooking methods are associated with inflammatory factors, renal function, and other hormones and nutritional biomarkers in older adults, *Sci. Rep.* 12(1) (2022) 16483. <https://doi.org/10.1038/s41598-022-19716-1>.
- [24] M. van Iterson, M.A. van de Wiel, J.M. Boer, et al., General power and sample size calculations for high-dimensional genomic data, *Stat. Appl. Genet. Mol. Biol.* 12(4) (2013) 449-467. <https://doi.org/10.1515/sagmb-2012-0046>.
- [25] M.Y. Schar, G. Corona, G. Soykan, et al., Excretion of avenanthramides, phenolic acids and their major metabolites following intake of oat bran, *Mol. Nutr. Food Res.* 62(2) (2018) 1700499. <https://doi.org/10.1002/mnfr.201700499>.
- [26] R. Hissong, K.R. Evans, C.R. Evans, Compound identification strategies in mass spectrometry-based metabolomics and pharmacometabolomics, *Handb Exp Pharmacol.* 277 (2023) 43-71. [https://doi.org/10.1007/164\\_2022\\_617](https://doi.org/10.1007/164_2022_617).
- [27] Y.Y. Li, T. Kind, J. Folz, et al., Spectral entropy outperforms MS/MS dot product similarity for small-molecule compound identification, *Nat. Methods.* 18(12) (2021) 1524-1531. <https://doi.org/10.1038/s41592-021-01331-z>.
- [28] S. Alseekh, A. Aharoni, Y. Brotman, et al., Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices, *Nat. Methods.* 18(7) (2021) 747-756. <https://doi.org/10.1038/s41592-021-01197-1>.
- [29] J.A. Westerhuis, E.J.J. van Velzen, H.C.J. Hoefsloot, et al., Multivariate paired data analysis: multilevel PLS-DA versus OPLS-DA, *Metabolomics.* 6(1) (2010) 119-128. <https://doi.org/10.1007/s11306-009-0185-z>.
- [30] F. Rohart, B. Gautier, A. Singh, et al., Mixomics: an R package for 'omics feature selection and multiple data integration, *PLoS Comput. Biol.* 13(11) (2017) e1005752. <https://doi.org/10.1371/journal.pcbi.1005752>.
- [31] L. Shi, J.A. Westerhuis, J. Rosen, et al., Variable selection and validation in multivariate modelling, *Bioinformatics.* 35(6) (2019) 972-980. <https://doi.org/10.1093/bioinformatics/bty710>.
- [32] F. Lindgren, B. Hansen, W. Karcher, et al., Model validation by permutation tests: applications to variable selection, *J. Chemom.* 10(5-6) (1996) 521-532. [https://doi.org/https://doi.org/10.1002/\(SICI\)1099-128X\(199609\)10:5/6<521::AID-CEM448>3.0.CO;2-J](https://doi.org/https://doi.org/10.1002/(SICI)1099-128X(199609)10:5/6<521::AID-CEM448>3.0.CO;2-J).
- [33] H.H. Jang, Y.M. Lee, J.S. Choe, et al., Validation of soy isoflavone intake and its health effects: a review of the development of exposure biomarkers, *Nutr. Res. Pract.* 15(1) (2021) 1-11. <https://doi.org/10.4162/nrp.2021.15.1.1>.
- [34] S. Hendrich, Bioavailability of isoflavones, *J. Chromatogr. B.* 777(1-2) (2002) 203-210. [https://doi.org/10.1016/S1570-0232\(02\)00347-1](https://doi.org/10.1016/S1570-0232(02)00347-1).
- [35] J.P. Zhang, Q.X. Guo, M.J. Wei, et al., Metabolite identification and pharmacokinetic profiling of isoflavones from black soybean in rats using ultrahigh-performance liquid chromatography with linear-ion-trap-orbitrap and triple-quadrupole tandem mass spectrometry, *J. Agric. Food. Chem.* 66(49) (2018) 12941-12952. <https://doi.org/10.1021/acs.jafc.8b04852>.
- [36] K. Zaheer, M. Humayoun Akhtar, An updated review of dietary isoflavones: nutrition, processing, bioavailability and impacts on human health, *Crit. Rev. Food. Sci. Nutr.* 57(6) (2017) 1280-1293. <https://doi.org/10.1080/10408398.2014.989958>.
- [37] M. Juarez Del Valle, J.E. Laino, A. de Moreno De LeBlanc, et al., Soyamilk fermented with riboflavin-producing *Lactobacillus plantarum* CRL 2130 reverts and prevents ariboflavinosis in murine models, *Br. J. Nutr.* 116(7) (2016) 1229-1235. <https://doi.org/10.1017/S0007114516003378>.
- [38] N. Suwannasom, I. Kao, A. Pruss, et al., Riboflavin: the health benefits of a forgotten natural vitamin, *Int. J. Mol. Sci.* 21(3) (2020) 950. <https://doi.org/10.3390/ijms21030950>.
- [39] H.M. Abdallah, M.A. Farag, M.M. Algandaby, et al., Osteoprotective activity and metabolite fingerprint via UPLC/MS and GC/MS of *lepidium sativum* in ovariectomized rats, *Nutrients.* 12(7) (2020) 2075. <https://doi.org/10.3390/nu12072075>.
- [40] Y. Fujimoto, M. Satoh, A new cytotoxic chlorine-containing polyacetylene from the callus of *panax ginseng*, *Chem. Pharm. Bull. (Tokyo).* 36(10) (1988) 4206-4208. <https://doi.org/10.1248/cpb.36.4206>.
- [41] S. Guan, K. Liu, Z.M. Liu, et al., UPLC-q-TOF/MS-based plasma and urine metabolomics contribute to the diagnosis of sepsis, *J. Proteome Res.* 21(1) (2022) 209-219. <https://doi.org/10.1021/acs.jproteome.1c00777>.
- [42] F. Saleem, M.P. Soos, Biotin deficiency, StatPearls, Treasure Island (FL), 2025.

- [43] J. Sorres, C. Nirma, V. Eparvier, et al., Pseudallicins a-d: four complex ovalicin derivatives from pseudallescheria boydii SNB-CN85, *Org. Lett.* 19(15) (2017) 3978-3981. <https://doi.org/10.1021/acs.orglett.7b01671>.
- [44] D. Amarakoon, W.J. Lee, G. Tamia, et al., Indole-3-carbinol: occurrence, health-beneficial properties, and cellular/molecular mechanisms, *Annu. Rev. Food Sci. Technol.* 14 (2023) 347-366. <https://doi.org/10.1146/annurev-food-060721-025531>.
- [45] N. Christinat, T. Stroheker, F. Badoud, et al., Naturally occurring indole-3-acetic acid in foods: a need for review of the current risk management measures in the EU, *Food Addit. Contam. Part A-Chem.* 42(4) (2025) 442-451. <https://doi.org/10.1080/19440049.2025.2459222>.
- [46] X.J. Li, B.B. Zhang, Y.Y. Hu, et al., New insights into gut-bacteria-derived indole and its derivatives in intestinal and liver diseases, *Front. Pharmacol.* 12 (2021) 769501. <https://doi.org/10.3389/fphar.2021.769501>.
- [47] A.M. Sheflin, E.C. Borresen, M.J. Wdowik, et al., Pilot dietary intervention with heat-stabilized rice bran modulates stool microbiota and metabolites in healthy adults, *Nutrients.* 7(2) (2015) 1282-1300. <https://doi.org/10.3390/nu7021282>.
- [48] M. Garcia-Aloy, R. Llorach, M. Urpi-Sarda, et al., Nutrimetabolomics fingerprinting to identify biomarkers of bread exposure in a free-living population from the PREDIMED study cohort, *Metabolomics.* 11(1) (2015) 155-165. <https://doi.org/10.1007/s11306-014-0682-6>.
- [49] J. Cheng, N.X. Wang, L. Yu, et al., Molecular insights into the one-carbon loss oxidation of indole-3-acetic acid, *ACS Catal.* 14(11) (2024) 8528-8540. <https://doi.org/10.1021/acscatal.4c02178>.
- [50] J. Tintelnor, Y. Xu, T.R. Lesker, et al., Microbiota-derived 3-IAA influences chemotherapy efficacy in pancreatic cancer, *Nature.* 615(7950) (2023) 168-174. <https://doi.org/10.1038/s41586-023-05728-y>.
- [51] M.J. Bender, A.C. McPherson, C.M. Phelps, et al., Dietary tryptophan metabolite released by intratumoral lactobacillus reuteri facilitates immune checkpoint inhibitor treatment, *Cell.* 186(9) (2023) 1846-1862. <https://doi.org/10.1016/j.cell.2023.03.011>.
- [52] L.P. Lin, D. Liu, J.C. Qian, et al., Post-ingestion conversion of dietary indoles into anticancer agents, *Natl. Sci. Rev.* 9(4) (2022) nwab144. <https://doi.org/10.1093/nsr/nwab144>.
- [53] V. Kriechbaumer, W.J. Park, A. Gierl, et al., Auxin biosynthesis in maize, *Plant biology (Stuttgart, Germany).* 8(3) (2006) 334-339. <https://doi.org/10.1055/s-2006-923883>.
- [54] J. Alberdi-Cedeño, M.L. Ibargoitia, M.D. Guillén, Bioactive compounds detected for the first time in corn oil: cyclic dipeptides and other nitrogenated compounds, *J. Food Compos. Anal.* 62 (2017) 197-204. <https://doi.org/https://doi.org/10.1016/j.jfca.2017.06.005>.
- [55] C. Favari, J.F. Rinaldi De Alvarenga, L. Sánchez-Martínez, et al., Factors driving the inter-individual variability in the metabolism and bioavailability of (poly)phenolic metabolites: a systematic review of human studies, *Redox Biol.* 71 (2024) 103095. <https://doi.org/10.1016/j.redox.2024.103095>.
- [56] K. Du, M.N. Zeng, Y.Z. Wang, et al., Two new beta-carboline alkaloids from the roots of anemone altaica, *J. Nat. Med.* 77(2) (2023) 397-402. <https://doi.org/10.1007/s11418-023-01679-w>.
- [57] D.C. Kim, T.H. Quang, C.S. Yoon, et al., Anti-neuroinflammatory activities of indole alkaloids from kanjang (korean fermented soy source) in lipopolysaccharide-induced BV2 microglial cells, *Food Chem.* 213 (2016) 69-75. <https://doi.org/10.1016/j.foodchem.2016.06.068>.
- [58] L.J. Yang, R.R. Li, Y.F. Cui, et al., Comparison of nutritional compositions of foxtail millet from the different cultivation regions by UPLC-q-orbitrap HRMS based metabolomics approach, *J. Food Biochem.* 45(10) (2021) e13940. <https://doi.org/10.1111/jfbc.13940>.
- [59] J. Walsh, J. Haddock, J.B. Blumberg, et al., Identification of methylated metabolites of oat avenanthramides in human plasma using UHPLC QToF-MS, *Int. J. Food Sci. Nutr.* 69(3) (2018) 377-383. <https://doi.org/10.1080/09637486.2017.1371116>.
- [60] P. Wang, J.L. Yang, A. Yerke, et al., Avenacosides: metabolism, and potential use as exposure biomarkers of oat intake, *Mol. Nutr. Food Res.* 61(7) (2017). <https://doi.org/10.1002/mnfr.201700196>.
- [61] S.M. Sang, Y.F. Chu, Whole grain oats, more than just a fiber: role of unique phytochemicals, *Mol. Nutr. Food Res.* 61(7) (2017). <https://doi.org/10.1002/mnfr.201600715>.
- [62] J.L. Yang, P. Wang, W.B. Wu, et al., Steroidal saponins in oat bran, *J. Agric. Food. Chem.* 64(7) (2016) 1549-1556. <https://doi.org/10.1021/acs.jafc.5b06071>.

- [63] L.O. Dragsted, Q. Gao, A. Scalbert, et al., Validation of biomarkers of food intake-critical assessment of candidate biomarkers, *Genes Nutr.* 13 (2018) 14. <https://doi.org/10.1186/s12263-018-0603-9>.