JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

pubs.acs.org/JAFC

Article

# Identifying Plasma and Urinary Biomarkers of Fermented Food Intake and Their Associations with Cardiometabolic Health in a Dutch Observational Cohort

Katherine J. Li, Kathryn J. Burton-Pimentel, Elske M. Brouwer-Brolsma, Carola Blaser, René Badertscher, Grégory Pimentel, Reto Portmann, Edith J. M. Feskens, and Guy Vergères*

Cite This: *J. Agric. Food Chem.* 2023, 71, 4426−4439

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** Identification of food intake biomarkers (FIBs) for fermented foods could help improve their dietary assessment and clarify their associations with cardiometabolic health. We aimed to identify novel FIBs for fermented foods in the plasma and urine metabolomes of 246 free-living Dutch adults using nontargeted LC−MS and GC−MS. Furthermore, associations between identified metabolites and several cardiometabolic risk factors were explored. In total, 37 metabolites were identified corresponding to the intakes of coffee, wine, and beer (none were identified for cocoa, bread, cheese, or yoghurt intake). While some of these metabolites appeared to originate from raw food (*e.g.*, niacin and trigonelline for coffee), others overlapped different fermented foods (*e.g.*, 4-hydroxybenzeneacetic acid for both wine and beer). In addition, several fermentation-dependent metabolites were identified (erythritol and citramalate). Associations between these identified metabolites with cardiometabolic parameters were weak and inconclusive. Further evaluation is warranted to confirm their relationships with cardiometabolic disease risk.

**KEYWORDS:** *biomarkers, cardiometabolic health, dietary assessment, fermented foods, metabolomics*



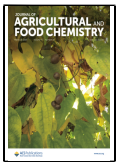
## INTRODUCTION

Accurate dietary assessment is crucial for detecting potential associations between diet and health. To date, many epidemiological studies still predominantly rely on self-reported dietary assessment methods, such as food frequency questionnaires (FFQ) and 24 h food recalls, which heavily depend on the memory and dedication of the participants.[1,2] As such, they are prone to multiple sources of measurement errors such as underreporting, inaccurate portion size estimation, and imprecision of food composition databases. Such measurement errors can reduce study power and miss detecting potential associations and may also lead to spurious findings.[3,4] Additionally, to capture the increasing diversity and complexity of modern diets, self-report methods require extensive food lists, which is burdensome for both participants and researchers. To address these limitations, food intake biomarkers (FIBs) have emerged as a more objective measure of dietary intake. Depending on their specificity, FIBs can be single compounds or a multimarker panel consisting of a combination of different compounds.[5] Recent advances in nutritional metabolomics have led to the identification of numerous candidate FIBs linked to the ingestion of a food, food group, or a dietary pattern.[3,6] However, FIBs for many foods in the diet have yet to be explored and validated—including fermented foods.

Fermented foods have been consumed since the beginning of human civilization and comprise up to 40% of the human diet.[7,8] The fermentation process not only improves the shelf life and organoleptic qualities of food, but it can also impart novel nutritional qualities that could improve human health.[9,10] A number of dietary intervention and epidemiological studies have suggested that the consumption of fermented foods positively affects cardiometabolic health, including weight maintenance, glucose metabolism, and cardiovascular health,[9,11−14] but the evidence is inconclusive. Thus, identification and validation of FIBs for fermented foods could improve the accuracy of dietary assessment and support further studies in obtaining more conclusive diet−health associations. Additionally, FIBs could also help elucidate the mechanisms of action that underpin the purported health benefits of fermented foods.

We previously conducted a systematic review of FIBs of fermented foods consumed worldwide and found several candidate FIBs at the food level, food group level, and/or fermentation level for several fermented foods, including wine, beer, bread, cocoa, coffee, postfermented tea, fermented soy, cheese, and yoghurt.[15] The majority of these FIBs were identified in postprandial studies with a small number of participants, and their relevance needs to be explored in free-living populations with complex, uncontrolled diets.[16] In the current work, we aimed to identify further FIBs of fermented foods consumed in The Netherlands by analyzing the plasma and urine metabolomes of a Dutch adult cohort using LC−MS

Received: August 12, 2022
Revised: February 12, 2023
Accepted: February 13, 2023
Published: February 28, 2023



and GC−MS. By utilizing a larger, free-living population, we expected the FIBs that emerge would be considered to be the most powerful and reliable indicators of habitual fermented food intake. In addition, we examined associations between the identified FIBs and several cardiometabolic risk parameters and composite risk scores.

## MATERIALS AND METHODS

**Study Population.** The Nutrition Questionnaires plus (NQplus) study is a prospective cohort study of 2048 Dutch men and women (20 to 70 years) with the aim to gather extensive data on participant demographics, anthropometrics, lifestyle, medical history, and cardiometabolic health outcomes.[17,18] Participants were recruited between June 2011 and February 2013. All measurements were performed according to a standardized protocol by trained research personnel. The study was approved by the ethical committee of Wageningen University and Research (protocol number NL34775.081.10) and conducted in agreement with the Declaration of Helsinki. Written informed consent was obtained from all participants prior to the start of the study.

Metabolomics analyses were performed on a subcohort of NQplus participants ($n$ = 531; $n$ = 485 with plasma samples and $n$ = 492 with urine samples) (herein referred to as the "metabolomics subcohort"). These participants were initially selected based on having a biosample collected within 14 days of completing either a FFQ or a 24 h recall. The FFQ was preferred over the 24 h recalls since it reflects more precisely the intake on any given day and is less sensitive to fluctuations in daily intake. Thus, for the selection of the most discriminant metabolites for identification, we focused the analyses on $n$ = 246 unique participants who had a biosample collected within 14 days of completing a FFQ ($n$ = 228 with plasma samples, and $n$ = 216 with urine samples) (herein referred to as the "identification subcohort", which is contained within the "metabolomics subcohort"). This criterion ensured that biosample collection occurred within the FFQ reference period of 1 month. To explore the stability of the FIBs with increasing time between biosample collection and FFQ completion, additional correlation analyses were conducted among participants with biosample collection within ±30 days ($n$ = 273), ±90 days ($n$ = 354), and ±180 days ($n$ = 501) of completing the FFQ, as well as within all 531 participants in the metabolomics subcohort.

**Food Frequency Questionnaire and Levels of Fermented Food Intake.** A detailed description of the validated, self-administered, semiquantitative 216-item FFQ used to assess habitual dietary intake has been reported previously.[17,18] In, participants completed the FFQ online and answered questions relating to frequency by selecting 1 of 10 frequency categories ranging from "never" to "6−7 days per week". Portion sizes were estimated using commonly used household measures. Total food intake (in g/d) was determined by multiplying consumption frequency by portion size as defined in the Dutch food composition tables (2011).[19] A total of 39 food items were classified as fermented, using criteria described previously[8] (Table S1). Most of the fermented foods and food groups in the FFQ have already been judged to have a good agreement with the intakes reported in 24 h recalls.[8] Only fermented foods and food groups that achieved "adequate" to "good" agreement in the validation study[8] (which is important for determining the reliability of self-reported intakes) and were consumed by at least a third of the population (which is important for the detection of potential FIBs in biosamples and selection of the most relevant FIBs) were included in the current analyses. These included fermented beverages (coffee, beer, and wine), fermented cereals/grains (white bread and whole-grain bread), fermented dairy (cheese and yoghurt), and cocoa-based products.

To facilitate the selection of FIBs that reflect the absolute dry weight of the different fermented foods considered within the fermented food groups (beverages, cereals/grains, cocoa-based products, and dairy), we further calculated the g dry matter/day intakes for each fermented food by subtracting the water weight of each food (in g/day) from the total intake (in g/day) (water weight determined from the Dutch food composition tables). Subsequently, energy adjustment was performed on all individual fermented foods as well as fermented food groups using the commonly used residual method.[20] All energy-adjusted fermented food intakes (in g/day and g dry matter/day) were then divided into tertiles representing the low (T1), mid (T2), and high (T3) levels of intake.

**Cardiometabolic Health Parameters.** Ten cardiometabolic health parameters collected at the baseline were included in the current analysis.[18] Height was determined using a stadiometer (SECA, Germany, nearest 0.1 cm), and weight was determined using a digital weighing scale (SECA, nearest 0.1 kg). The BMI was calculated by dividing weight (in kg) by height (in $m^2$). Waist circumference was measured twice using a nonflexible measuring tape (SECA 201, nearest 0.5 cm) and averaged. Enzymatic methods[21] were applied to assess fasting plasma glucose, total cholesterol, HDL cholesterol, and serum triglycerides using a Dimension Vista 1500 automated analyzer (Siemens, Erlangen, Germany) or Roche Modular P800 chemistry analyzer (Roche Diagnostics, Indianapolis, USA). Plasma LDL cholesterol was calculated with the Friedewald equation.[22] Hemoglobin A1c (HbA1c) concentrations in whole blood were determined by HPLC using an ADAMS A1c HA-8160 analyzer (A. Menarini Diagnostics). Systolic and diastolic blood pressure were measured using a digital blood pressure monitor (IntelliSense HEM-907, Omron Healthcare, USA); the first measurement was discarded, and the remaining (up to 6) measurements were averaged. Participants were classified as having hypertension, suboptimal cholesterol, or type II diabetes based on the cutoffs and definitions described in relevant guidelines of the European Society of Cardiology/European Atherosclerosis Society (ESC/EAS)[23−25] and having metabolic syndrome based on the harmonized guidelines of the International Diabetes Federation (IDF) et al.[26]

Two composite risk scores were also determined as previously described,[27] consisting of a continuous metabolic syndrome (MetS) score using summed age- and sex-adjusted standardized residuals ($z$-scores) of individual MetS parameters[28−30] and the European Systematic COronary Risk Evaluation (SCORE)[31,32] evaluating 10 year risk of fatal cardiovascular disease.

**Covariates.** All covariate data relevant to the current work (age, sex, education level, smoking status, physical activity, alcohol consumption, and dietary intake) were collected via questionnaires.[18] For educational level, participants with no education or primary/lower vocational education were classified under "low", participants who completed lower secondary or intermediate vocational education were classified under "intermediate", and participants who completed higher secondary or higher vocational education, or university, were classified under "high". A "smoker" was defined as a current smoker and former smoker who quit >35 years old and a "nonsmoker" as never smoker and former smoker who quit <35 years old.[33] Information on the participants' usual physical activity over the past 4 weeks was obtained using the validated Activity Questionnaire for Adults and Adolescents (AQuAA), which provides the time spent on sedentary-, light-, moderate-, and vigorous-intensity activities in min/week.[34] Intake levels of alcohol and different foods were assessed by a FFQ, as described above.

**LC−MS Metabolomics Analysis.** EDTA plasma and 24 h urine samples collected for NQplus were used for metabolomics analyses. All samples were thawed on ice and kept at 4 °C during analysis. Prior to LC−MS analysis, phospholipids were removed from plasma samples to limit ion suppression using a Phree filter (Phenomenex Inc., Torrance, CA). Urine samples were normalized based on the specific gravity as determined by the refractive index (refractometer RE40, Mettler Toledo, Switzerland), as described previously.[35,36] LC−MS metabolomics analysis was performed using an UltiMate 3000 RS UPLC system (Thermo Fisher Scientific, Waltham, MA) with a Waters Acquity UPLC HSS T3 column (length 150 mm, diameter 2.1 mm, and particle size 1.8 μm), coupled with a maXis 4G + quadrupole time-of-flight mass spectrometer (Bruker Daltonik GmbH, Bremen, Germany). A gradient was run from 5% to 95% of mobile phase A within 15 min at 0.4 mL/min. Mobile phase A



https://doi.org/10.1021/acs.jafc.2c05669
J. Agric. Food Chem. 2023, 71, 4426−4439

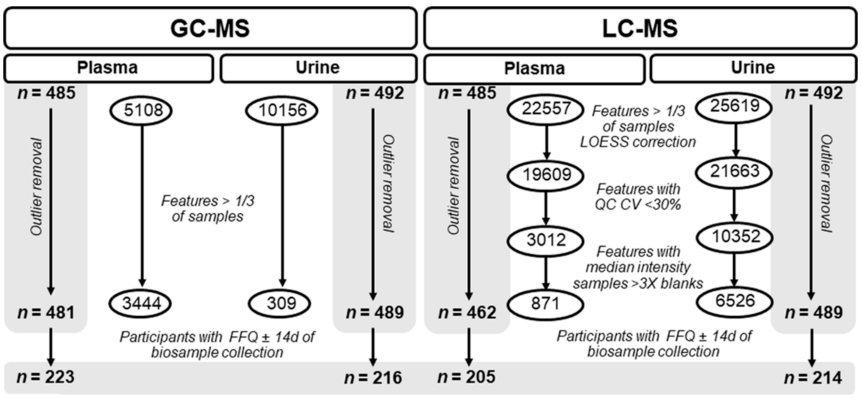



**Figure 1.** GC−MS and LC−MS metabolomics workflow for feature filtration and outlier removal.

consisted of Milli-Q water with 0.1% formic acid, and mobile phase B consisted of acetonitrile with 0.1% formic acid. The column was heated to 35 °C with a postcolumn cooler set to 25 °C. The resulting system pressure was ∼600 bar, dependent on the actual composition of the mobile phase at the specific time. The mass spectrometer ESI was operated in positive ion mode, and spectra were recorded from 75 to 1500 $m/z$. Collision-induced dissociation was performed using energies from 20 to 70 eV. 5 $\mu$L of filtered plasma or normalized urine from each sample was injected once in a randomized sequence. Quality control (QC) pools were prepared from plasma or urine samples by mixing all samples of each sample type at equal volumes. QC samples were injected at five sample intervals for signal drift correction. Blanks (consisting of ultrafiltered LC−MS-grade water) were also injected at the beginning and end of each batch for the detection of contaminants.

Progenesis QI (v.2.3.6198.24128, NonLinear Dynamics Ltd., Newcastle upon Tyne, United Kingdom) was used for retention time correction, peak-picking, deconvolution, adducts annotation, and normalization (default automatic sensitivity and without minimum peak width). All solvents and reagents were purchased from Sigma-Aldrich Chemie GmbH (Buchs, Switzerland).

**GC−MS Metabolomics Analysis.** Plasma and urine samples were prepared for GC−MS analysis as described previously for serum[37] and urine.[38] Specifically, for each 100 $\mu$L plasma sample, 50 $\mu$L of an internal standard solution (labeled D-sucrose, 13C12, 98%, Cambridge Isotope Laboratories, Inc., Cambridge, UK, $c \approx 0.16$ mg/mL in water) was added, followed by precipitation with 300 $\mu$L of cold methanol, centrifugation, transfer of supernatant (370 $\mu$L), and drying using a vacuum centrifuge. Urine samples were normalized prior to analysis using the refractive index methods described above for the LC−MS analysis. For each 100 $\mu$L urine sample, 50 $\mu$L of an internal standard solution (labeled D-sucrose) was added and dried using a vacuum centrifuge. The plasma and urine samples further underwent a two-step derivatization (methoximation with *O*-methylhydroxylamine hydrochloride followed by silylation with *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA)) and were analyzed by GC−MS 7890B/MS5977A (Agilent Technologies, Santa Clara, CA, US) with a CombiPAL autosampler (CTC-Analytics AG, Zwingen, Switzerland) and a DB-5 ms fused silica capillary column (60 m, 0.25 mm i.d., 0.25 $\mu$m film thickness, Agilent Technologies, Basel, Switzerland). The samples were injected using a multimode injector according to the following temperature program: initially 90 °C, a heating rate of 900 °C/min until 280 °C, held for 5 min and cooled at a rate of −30 °C/min, and maintained at 250 °C. 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 36 min, equilibration time 1 min. The MS detection mass ranged from 28.5 to 600 Da, the MS source temperature was 230 °C, and the MS Quad temperature was 150 °C. Electron ionization was performed with 70 eV. Each batch was initiated by three injections of QC samples for equilibration and after every fifth plasma sample, a fresh QC was injected. At the start and end of each batch, a blank sample (Milli-Q water) was included. QC samples and blank samples underwent the same sample preparation as plasma samples.

Agilent data files acquired from GC−MS analysis were deconvoluted and converted into CEF files using Agilent MasshunterProfinder (Agilent Technologies, Santa Clara, US). Data files were further processed in Agilent Mass Profiler Professional (Agilent Technologies, Santa Clara, U.S.) to perform alignment and compound identification. Features with retention time before 10 min (reagents region) were removed. 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, the suggested markers were analyzed in MassHunter Quantitative Analysis (Agilent Technologies, Santa Clara, US). 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 internal standard [labeled D-sucrose (ion 220)].

**Metabolomics Data Preprocessing.** The dataset was corrected to account for signal drift and reduced via multiple filtering steps to remove features with poor repeatability and potential contaminants (Figure 1). Principal component analyses (PCAs) of the QCs for both LC−MS and GC−MS present the relative stability of the analysis (Figure S1). For LC−MS, the QC-based robust locally estimated scatterplot smoothing signal correction method was applied for signal drift correction[39] using R (v.3.6.3).[40] Features resulting from LC−MS analysis were removed if they had poor repeatability (detected in less than one-third of samples), a relative standard deviation > 30% in the QC samples, and a median in the QC samples that was <3 times higher than the median calculated for the blanks. For GC−MS, features detected in less than one-third of samples were removed (features that had high levels in blanks or originated from the GC column were removed after identification to ensure all features captured during automatic detection are retained and further inspected for relevance). Exploratory analyses were performed, and metabolomics sample outliers, defined as observations clearly falling outside Hotelling's T2 tolerance eclipse (95% confidence interval) in the PCA score plot, were identified and excluded ($n$ = 23 LC−MS plasma, $n$ = 3 LC−MS urine, and $n$ = 4 GC−MS plasma, $n$ = 3 GC−MS urine) (Figure S2).

**Selection of Discriminant Metabolites by Univariate and Multivariate Statistics.** We performed several complementary univariate and multivariate statistical tests to select and confirm the most consistent features to proceed with metabolite identification.[41] Differences in levels of features by tertiles of intake for fermented foods and groups (T1, T2, and T3) were assessed by a Kruskal−Wallis test followed by a posthoc Conover-Iman pairwise comparison test. An additional step was conducted to select features with higher levels in higher tertiles compared to lower tertiles (*i.e.*, a median of T3 > T1, T3 > T2, and T2 > T1). To determine the strength and direction of the associations between fermented food intakes and features, nonparameteric Spearman's rank correlation coefficients ($r_s$)


https://doi.org/10.1021/acs.jafc.2c05669
J. Agric. Food Chem. 2023, 71, 4426−4439

**Table 1. Characteristics of the Study Population**[a]

| characteristic | metabolomics subcohort ($n$ = 531) | identification subcohort ($n$ = 246) | $p$-value |
|---|---|---|---|
| | Demographics | | |
| age, years | 57 (46−63) | 58 (46−65) | 0.26 |
| education, $n$ (%) | | | 0.73 |
| low | 37 (7) | 19 (8) | |
| intermediate | 148 (28) | 77 (31) | |
| high | 344 (65) | 149 (61) | |
| smoking status, $n$ (%) | | | 0.20 |
| smoker | 118 (26) | 70 (31) | |
| nonsmoker | 343 (74) | 159 (69) | |
| physical activity, min/week | 2136 ± 1093 | 2043 ± 1046 | 0.37 |
| supplement use, $n$ (%) | 0.8 ± 1.2 | 0.7 ± 1.2 | 0.58 |
| | Dietary Factors | | |
| total energy intake, kcal/day | 2128 ± 499 | 2220 ± 530 | **0.02*** |
| | Macronutrients | | |
| fat, g/day (En %) | 84 ± 25 (36%) | 90 ± 27 (36%) | **0.01*** |
| carbohydrates, g/day (En %) | 230 ± 60 (43%) | 237 ± 63 (43%) | 0.17 |
| protein, g/day (En %) | 77 ± 18 (14%) | 80 ± 18 (14%) | 0.06 |
| fiber, g/day | 25 ± 7 | 25 ± 7 | 0.80 |
| sodium, mg/day | 2261 ± 653 | 2375 ± 711 | **0.03*** |
| | Fermented Foods and Groups | | |
| total fermented beverages, g/day | 592 (324−799) | 629 (406−865) | 0.26 |
| coffee, g/day | 406 (174−638) | 406 (196−638) | 0.48 |
| wine, g/day | 25 (4−87) | 20 (0−80) | 0.31 |
| beer, g/day | 9 (0−79) | 20 (0−118) | **0.04*** |
| total fermented cereals/grains, g/day | 130 (88−166) | 133 (88−170) | 0.41 |
| whole-grain bread, g/day | 80 (47−112) | 77 (41−114) | 0.51 |
| white bread, g/day | 2 (0−8) | 2 (0−10) | 0.24 |
| cocoa, g/day | 4 (1−8) | 4 (1−8) | 0.80 |
| total fermented dairy, g/day | 152 (76−245) | 151 (69−240) | 0.79 |
| cheese, g/day | 25 (13−42) | 28 (14−46) | 0.24 |
| yoghurt, g/day | 89 (29−139) | 82 (21−139) | 0.30 |
| | Other Foods and Groups | | |
| tea, g/day | 174 (67−406) | 174 (67−406) | **0.04*** |
| alcoholic drinks, g/day | 81 (18−207) | 108 (19−245) | 0.26 |
| soft drinks, g/day | 5 (0−42) | 13 (0−54) | **0.04*** |
| fruits, g/day | 217 (86−238) | 166 (81−233) | 0.10 |
| vegetables, g/day | 150 (97−204) | 140 (94−196) | 0.11 |
| potatoes, g/day | 67 (37−87) | 67 (37−87) | 0.29 |
| legumes, g/day | 38 (19−79) | 38 (22−79) | 0.89 |
| meat products, g/day | 72 (46−98) | 79 (54−105) | 0.053 |
| eggs and egg products, g/day | 9 (7−18) | 14 (7−18) | **0.03*** |
| fish, g/day | 11 (6−16) | 11 (6−16) | 0.72 |
| nuts and seeds, g/day | 13 (6−25) | 13 (6−26) | 0.71 |
| sauces, spreads, and cooking fats, g/day | 41 (28−54) | 42 (30−57) | 0.21 |
| salty and processed snack foods, g/day | 35 (16−59) | 37 (20−64) | 0.16 |
| sugary confectionary and desserts, g/day | 70 (47−104) | 78 (50−113) | 0.09 |
| | Cardiometabolic Factors | | |
| BMI, kg/m$^2$ | 25.1 (22.9−27.2) | 25.5 (23.2−28.0) | 0.12 |
| BMI category, $n$ (%) | | | 0.61 |
| underweight (<18.5 kg/m$^2$) | 4 (1) | 2 (1) | |
| normal weight (18.5−24.9 kg/m$^2$) | 249 (7) | 103 (42) | |
| overweight or obese (≥25−29.9 kg/m$^2$) | 278 (52) | 141 (57) | |
| waist circumference, cm | 91 ± 12 | 93 ± 12 | **0.04*** |
| diastolic blood pressure, mm Hg | 73.7 ± 10.4 | 74.5 ± 10.8 | 0.38 |
| systolic blood pressure, mm Hg | 125.5 ± 16.0 | 128.7 ± 16.6 | **0.01*** |
| hypertension, $n$ (%) | | | 0.42 |
| hypertension[b] | 109 (20.6) | 62 (25.2) | |
| normal or optimal | 421 (79.4) | 184 (74.8) | |
| hypertension treatment, $n$ (%) | | | 0.95 |
| being treated with medication and/or diet | 69 (13.0) | 36 (14.6) | |
| not being treated | 462 (87.0) | 210 (85.4) | |



https://doi.org/10.1021/acs.jafc.2c05669
J. Agric. Food Chem. 2023, 71, 4426−4439

**Table 1. continued**

| characteristic | metabolomics subcohort ($n$ = 531) | identification subcohort ($n$ = 246) | $p$-value |
|---|---|---|---|
| | Cardiometabolic Factors | | |
| plasma total cholesterol, mmol/L | 5.4 ± 1.0 | 5.3 ± 1.0 | 0.15 |
| plasma LDL cholesterol, mmol/L | 3.3 ± 0.9 | 3.2 ± 0.9 | 0.63 |
| plasma HDL cholesterol, mmol/L | 1.6 ± 0.4 | 1.5 ± 0.4 | **0.01*** |
| serum triglycerides, mmol/L | 1.0 (0.7−1.4) | 1.0 (0.7−1.4) | 0.74 |
| suboptimal cholesterol, $n$ (%) | 398 (75.0) | 182 (74.0) | 0.84 |
| high cholesterol treatment, $n$ (%) | | | 0.76 |
| being treated with medication and/or diet | 56 (10.5) | 23 (9.3) | |
| not being treated | 475 (89.5) | 223 (90.7) | |
| HbA1c, mmol/mol | 35.5 (34.0−38.0) | 35.8 (34.0−38.0) | 0.97 |
| fasting glucose, mmol/L | 5.4 (5.1−5.8) | 5.3 (5.0−5.8) | 0.11 |
| diabetes, $n$ (%) | 13 (2.4) | 6 (2.4) | 0.98 |
| diabetes treatment, $n$ (%) | | | 0.77 |
| being treated with medication and/or diet | 15 (2.8) | 5 (2.0) | |
| not being treated | 516 (97.2) | 241 (98.0) | |
| metabolic syndrome, $n$ (%) | 67 (12.6) | 33 (13.4) | 0.85 |
| SCORE, $n$ (%) | | | 0.25 |
| ≥15% | 6 (1.3) | 6 (2.6) | |
| 10−14% | 17 (3.7) | 14 (6.2) | |
| 5−9% | 73 (16.0) | 42 (18.5) | |
| 1−4% | 211 (46.2) | 98 (43.2) | |
| <1% | 150 (32.8) | 67 (29.5) | |

[a]BMI, body mass index; HDL, high-density lipoprotein; LDL, low-density lipoprotein; SCORE, Systematic COronary Risk Evaluation; SD, standard deviation. Values are presented as mean ± SD unless otherwise specified. Missing values for the metabolomics subcohort: education ($n$ = 2), smoking ($n$ = 70), physical activity ($n$ = 296), LDL cholesterol ($n$ = 4), HDL cholesterol ($n$ = 4), Hb1Ac ($n$ = 5), glucose ($n$ = 4), SCORE ($n$ = 74). Missing values for the identification subcohort: education ($n$ = 1), smoking ($n$ = 17), physical activity ($n$ = 41), LDL-cholesterol ($n$ = 3), HDL-cholesterol ($n$ = 3), Hb1Ac ($n$ = 4), glucose ($n$ = 3), SCORE ($n$ = 19). Differences between the metabolomics and identification subcohorts were assessed using the $t$-test (for normally distributed continuous variables), Wilcoxon test (for skewed continuous variables), or $\chi$-squared test (for categorical variables). Significant p-values are bolded and indicated by an asterisk (*). [b]Inclusive of grade 1 hypertension, grade 2 hypertension, and isolated systolic hypertension.

were calculated; significant correlations with $r_s > 0.20$ were selected for further analysis. For all univariate statistical tests, $p$-values were adjusted for the false discovery rate (FDR) using the method of Benjamini and Hochberg,[42] and FDR-adjusted $p \leq 0.05$ was set as the significance threshold.

Two multivariate tests were also conducted to further unveil and confirm features that discriminate between tertiles of fermented food intake. Partial least-square discriminant analysis (PLS-DA) was performed to identify features that differentiate the lowest and highest tertiles of intake for each fermented food or food group (SIMCA-P software v.15.0; Umetrics). The dataset was scaled using the unit variance (UV) method. The quality and validity of the models were evaluated by the goodness-of-fit parameter (R2Y > 0.5), the predictive ability parameter (Q2 > 0.2), and permutation tests with 999 random permutations to exclude any random separation of the sample groups.[43] Permutation plots (correlations of the observed and permuted data on the $X$-axis against the R2Y and Q2 on the $Y$-axis) for all models were visually interpreted as follows: the model was considered to be well guarded against overfitting if the Q2 values of the permuted dataset were lower than the Q2 value of the actual dataset to the observed dataset. Finally, the most discriminant features from these models were selected based on variable importance in projection (VIP) scores (VIP > 1 as a cutoff value). Second, we used random forests to model these data and further select the most discriminant features between T1, T2, and T3 of fermented food or food group intake, using the randomForest package.[44] The dataset was split into training (0.75) and test (0.25) datasets. For tuning the random forest, the number of trees ranged from 500 to 800 and the node sizes from 1 to 10. The "mtry" parameter was set to $x$ (0.01, 0.05, 0.15, 0.25, 0.333, and 0.4), where $x$ is the number of features considered for the model. We then implemented a full Cartesian grid search to choose the best model using the out-of-bag estimates generated from the random forest model. The results and variable importance from this step were further subjected to permutation testing using the "altmann" method ($n$ = 500)[45] applied in the ranger package.[46] For features selected from multivariate analysis, the Wilcoxon test (for two comparisons) or Kruskal−Wallis test followed by a posthoc Conover-Iman pairwise comparison test (for three comparisons) was also conducted (non-FDR-adjusted $p \leq 0.05$) as a separate validation test of the features selected from these models.

Given the large number of significant features revealed across complementary univariate and multivariate tests (586 plasma and 151 urinary metabolites from GC−MS and 110 plasma and 4473 urinary metabolites from LC−MS; data not shown in tables), those significant in at least two of the four statistical tests were prioritized and selected for identification. For urinary features measured by LC−MS, a large number of features remained significant; thus, an additional criterion of Spearman's FDR $p$-value $\leq 1 \times 10^{-10}$ had to be applied to select a number of features that could feasibly be identified. A summary of the significant features across at least two of the four statistical tests (and prioritized for identification) are provided in Table S2. Aside from PLS-DA, all analyses were performed in R (v.3.6.3).[40]

**Metabolite Identification.** For LC−MS, the Human Metabolome Database,[47] the MassBank of North America,[48] the National Institute of Standards and Technology database (NIST v17), and METLIN[49] were used to screen the identity of metabolites with a 10 ppm mass accuracy threshold. Identity suggestions from databases were then screened based on the chemical and biological relevance of each suggested metabolite identification (as provided on HMDB and/or through a search of the compound name on PubMed and Google) and confirmed by MS fragmentation data (where available). Pure analytical standards were then purchased for the tentatively identified and most biologically plausible compounds and injected at two concentrations in sample QCs and in the solvent. For GC−MS, the Golm Metabolome Database[50] and NIST v17 were used to screen the identity of compounds, and an internal database of internal standards



https://doi.org/10.1021/acs.jafc.2c05669
J. Agric. Food Chem. 2023, 71, 4426−4439

Table 2. Tertiles of Fermented Food Intake in the Identification Subcohort ($n$ = 246)[a]

| | energy-adjusted intakes (g/day) | | | | energy-adjusted intakes (g dry matter/day) | | | |
|---|---|---|---|---|---|---|---|---|
| food group | mean ± SD | T1 ($n$ = 82) | T2 ($n$ = 82) | T3 ($n$ = 82) | mean ± SD | T1 ($n$ = 82) | T2 ($n$ = 82) | T3 ($n$ = 82) |
| total FB | 638 ± 398 | 264 (124, 378) | 615 (551, 679) | 978 (857, 1205) | 22 ± 20 | 6 (3, 8) | 17 (14, 21) | 37 (30, 51) |
| coffee | 466 ± 297 | 142 (63, 238) | 453 (418, 510) | 691 (640, 903) | 5 ± 3 | 1 (1, 2) | 5 (4, 5) | 7 (6, 9) |
| beer | 112 ± 202 | −14 (−32, 11) | 48 (37, 71) | 208 (136, 374) | 9 ± 16 | −1 (−3, 1) | 4 (3, 6) | 17 (11, 30) |
| wine | 61 ± 89 | 3 (−7, 7) | 25 (16, 38) | 130 (92, 187) | 8 ± 12 | 0 (−1, 1) | 4 (2, 5) | 17 (12, 27) |
| total FCG | 134 ± 60 | 79 (61, 94) | 132 (115, 144) | 182 (168, 219) | 85 ± 37 | 52 (38, 61) | 85 (74, 91) | 116 (105, 137) |
| white bread | 8 ± 13 | −1 (−2, 1) | 4 (3, 5) | 17 (10, 30) | 5 ± 8 | 0 (−1, 1) | 3 (2, 3) | 11 (7, 19) |
| whole-grain bread | 82 ± 56 | 29 (12, 42) | 77 (68, 88) | 131 (114, 149) | 52 ± 35 | 20 (8, 27) | 49 (43, 55) | 83 (72, 93) |
| cocoa | 6 ± 9 | 1 (0, 1) | 3 (3, 4) | 10 (8, 17) | 6 ± 9 | 1 (0, 1) | 3 (2, 4) | 10 (8, 17) |
| tTotal FD | 170 ± 121 | 55 (33, 69) | 142 (126, 173) | 286 (238, 342) | 35 ± 19 | 16 (11, 21) | 32 (28, 38) | 54 (47, 63) |
| cheese | 34 ± 26 | 12 (7, 17) | 27 (23, 32) | 56 (47, 73) | 19 ± 15 | 7 (4, 10) | 16 (13, 18) | 31 (27, 42) |
| yoghurt | 89 ± 79 | 7 (0, 21) | 82 (59, 96) | 139 (139, 193) | 11 ± 10 | 1 (0, 3) | 11 (8, 13) | 20 (17, 24) |

[a]FB, fermented beverages; FCG, fermented cereals and grains; FD, fermented dairy. Values are reported as median (IQR) unless otherwise specified.

was used to confirm the metabolite identification. In the case that stereoisomeric forms of selected discriminating features were identified, the peak with a higher response was further evaluated. The list of standard suppliers is provided in Table S3. For both LC−MS and GC−MS, the level of identification of each discriminant metabolite is defined according to the Metabolomics Standards Initiative (MSI) recommendations,[51] as follows: level 1, compounds identified by comparison to a pure reference standard based on spectral data (LC: molecular weight with a 10 ppm accuracy threshold, fragmentation pattern when available, isotopic distribution, and retention time with a 10% accuracy threshold; GC: based on spectral data and retention indices (RIs) with a 5% accuracy threshold and 10% for very large peaks); level 2, based on spectral data but without chemical standards (LC: fragmentation pattern match to library spectral data of at least two major peaks; GC: library match factor >80%); level 3, putatively characterized compound classes; and level 4, unknown compound. Details of the identification features of metabolites analyzed from GC−MS (37 plasma and 75 urinary metabolites) and LC−MS (13 plasma and 89 urinary metabolites) are presented in Tables S4 and S5, respectively. The metabolites corresponded to the intakes of total fermented beverages (FBs) (number of metabolites: 112), wine (89), coffee (72), beer (17), white bread (9), total fermented cereals/grains (FCG) (1), total fermented dairy (FD) (1), cheese (1), and cocoa (1) (none for whole-grain bread or yoghurt).

**Associations between Identified Metabolites and CMD Risk Parameters.** Participant characteristics are reported as number (percentages), mean (standard deviation) for normally distributed variables, or medians (interquartile range) for skewed variables. Multivariable adjusted linear regression and restricted cubic spline regression were used to evaluate the associations between the identified metabolites and CMD risk factors. CMD risk parameters acting as dependent variables that were not normally distributed were log-transformed, which included: BMI, plasma HbA1c, plasma glucose, serum triglycerides, and SCORE. All variables were normalized by $z$-scores prior to analysis to allow comparability across associations. Analyses were performed unadjusted (model 0), adjusted for age (years) and sex (male, female) (model 1) + physical activity (minutes/week), smoking (smoker/non-smoker), and education level (high, intermediate, low) (model 2) + dietary factors (g/day) (model 3). For associations with continuous MetS, which already takes into account age and sex, analyses were performed unadjusted (model 0) and fully adjusted for smoking, physical activity, education, and dietary factors (model 3). For associations with SCORE, which already takes into account age, sex, and smoking status, analyses were performed unadjusted (model 0) and fully adjusted for physical activity, education, and dietary factors (model 3). Dietary factors included in the fully adjusted models included those indicated in the literature to be important for CMD risk in addition to those significantly correlated with the identified metabolites and included vegetables, fruits, alcohol, meat, and confectionary/desserts. All analyses were performed in R (Version 3.6.3).[52] For all models, the level of significance was set at $p \leq 0.05$. To account for multiple comparisons, FDR-adjusted $p$-values are also presented.

## RESULTS

**Characteristics of the Population.** The characteristics of the participants in the metabolomics and identification subcohorts are presented in Table 1. The median age of the participants was ∼58 years, and the majority were highly educated (>60%) and nonsmokers (>69%). No significant differences were observed in background demographics between the two subcohorts. Among the dietary factors, participants in the identification subcohort had significantly higher intakes of total energy, fat, sodium, beer, soft drinks, and egg products compared to participants in the metabolomics subcohort but with a similar interquartile range (significant differences were also observed for tea intake but medians were comparable) ($p \leq 0.05$). Among cardiometabolic parameters, participants in the identification subcohort have a slightly larger waist circumference, higher systolic blood pressure, and lower plasma HDL-cholesterol than participants in the metabolomics subcohort. However, although significant, the differences observed are relatively minor and do not pertain to the broader indicators of health linked to each measure (*e.g.*, BMI, hypertension, and suboptimal cholesterol). The distribution of participant risk for continuous MetS and SCORE is presented in Figure S3.

**Intake Levels of Different Fermented Foods.** The levels of intake of fermented foods in the identification subcohort (mean and tertiles) are presented in Table 2 (in both absolute g/day and g dry matter/day). Out of the fermented food groups evaluated, the highest intake on a g/day basis was total FB followed by total FD, while the highest mean intake of foods on a g dry/matter per day was total FCG. Out of individual fermented foods, coffee had the highest intake among all other fermented foods on a g/day basis (466 g/day) but the lowest intake on a g dry matter/day basis (similar trends were observed for wine and beer). Conversely, intakes of cocoa remained the same regardless of g/day or g dry matter/day (similar trends were observed for white and whole-grain bread, and cheese).

**Biomarkers Identified for Fermented Food Intake.** A total of 12 plasma metabolites and 27 urinary metabolites were identified. An overview of the candidate FIBs identified for



https://doi.org/10.1021/acs.jafc.2c05669
J. Agric. Food Chem. 2023, 71, 4426−4439

**Table 3. Overview of Identified Plasma Metabolites Discriminant for the Intake of Various Fermented Foods**

| | | | | Fermented food or food group | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Identification | HMDB ID | Platform | ID level | Total FB | Coffee | Wine | Beer | Total FCG | WG bread | WT bread | Cocoa | Total FD | Cheese | Yoghurt |
| Erythritol | HMDB0002994 | GC-MS | 1 | | | ● | | | | | | | | |
| 2-Hydroxybutyric acid | HMDB0000008 | GC-MS | 1 | ● | | | | | | | | | | |
| L-Cysteine | HMDB0000574 | GC-MS | 1 | ● | | | ● | | | | | | | |
| Dodecanoic acid[a] | HMDB0000638 | GC-MS | 1 | | | | | | | ● | | | | |
| Xylitol | HMDB0242149 | GC-MS | 1 | ● | | ● | | | | | | | | |
| trans-Aconitic acid | HMDB0000958 | GC-MS | 1 | ● | | | | | | | | | | |
| Quinate | HMDB0003072 | GC-MS | 1 | ● | ● | | | | | | | | | |
| L-Phenylalanine | HMDB0000159 | GC-MS | 1 | ● | | | | | | | | | | |
| Isoleucine | HMDB0000172 | GC-MS | 1 | ● | | | | | | | | | | |
| Glutamic acid | HMDB0000148 | LC-MS | 1[b] | | ● | | | | | | | | | |
| Trigonelline | HMDB0000875 | LC-MS | 1[b] | | ● | | | | | | | | | |
| Hydroxy(*iso*)butyric acid | HMDB0000023 | LC-MS | 2 | | | ● | | | | | | | | |

[a]FB, fermented beverages; FCG, fermented cereals and grains; FD, fermented dairy; GC−MS, gas chromatography−mass spectrometry; ID, identification; LC−MS, liquid chromatography−mass spectrometry; WG, whole grain; WT, white. Based on Spearman's correlations this metabolite is negatively associated with the fermented foods and food groups indicated but discriminant based on statistical significance in other tests (*e.g.*, PLS-DA, random forest). [b]Matched on retention time and mass; information on fragmentation not available.

**Table 4. Overview of Identified Urine Metabolites Discriminant for the Intake of Various Fermented Foods**

| | | | | Fermented food or food group | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Identification | HMDB ID | Platform | ID level | Total FB | Coffee | Wine | Beer | Total FCG | WG bread | WT bread | Cocoa | Total FD | Cheese | Yoghurt |
| D-Psicose[a] | HMDB0250793 | GC-MS | 1 | ● | ● | | ● | | | | | | | |
| Glycine[a] | HMDB0000123 | GC-MS | 1 | ● | | | | | | | | | | |
| D-Gluconate[a] | HMDB0000625 | GC-MS | 1 | | | ● | | | | | | | | |
| Guaiacol | HMDB0001398 | GC-MS | 1 | ● | | | | | | | | | | |
| 3-deoxy-D-ribo-hexonic acid gamma-lactone | Not available | GC-MS | 2 | ● | ● | | | | | | | | | |
| 2,3-Dihydroxypropyl phosphoric acid | Not available | GC-MS | 2 | ● | | | ● | | | | | | | |
| Glucuronic acid | HMDB0000127 | GC-MS | 2 | | | ● | | | | | | | | |
| m-Cresol[a] | HMDB0002048 | GC-MS | 2 | ● | | | | | | | | | | |
| 4-Hydroxybenzeneacetic acid | HMDB0000020 | GC-MS | 2 | ● | | ● | ● | | | | | | | |
| 3-Hydroxyhippuric acid | Not available | GC-MS | 2 | ● | | | | | | | | | | |
| D-Lactose | HMDB0041627 | GC-MS | 1 | ● | | | ● | | | | | | | |
| Niacin (Nicotinate/Vitamin B3) | HMDB0001488 | GC-MS | 1 | ● | ● | | | | | | | | | |
| Catechol | HMDB0000957 | GC-MS | 1 | ● | ● | | | | | | | | | |
| Citramalate | HMDB0000426 | GC-MS | 1 | | | ● | | | | | | | | |
| Erythritol | HMDB0002994 | GC-MS | 1 | ● | | ● | | | | | | | | |
| Tartaric acid | HMDB0000956 | GC-MS | 1 | ● | | ● | | | | | | | | |
| 3,4-Dihydroxyhydrocinnamic acid | HMDB0000423 | GC-MS | 2 | ● | | | | | | | | | | |
| D-Fucitol[a] | HMDB0304954 | GC-MS | 2 | | ● | | | | | | | | | |
| 2,3-Dihydroxybutanoic acid | HMDB0245394 | GC-MS | 2 | ● | | ● | ● | | | | | | | |
| Arabinofuranose | HMDB0012325 | GC-MS | 2 | ● | | ● | | | | | | | | |
| Furoylglycine | HMDB0000439 | GC-MS | 2 | ● | ● | | | | | | | | | |
| 2-Keto-l-gluconic acid[a] | HMDB0245186 | GC-MS | 2 | ● | ● | | | | | | | | | |
| Ethyl α-D-glucopyranoside | HMDB0252035 | GC-MS | 2 | ● | | ● | ● | | | | | | | |
| Glyceryl-glycoside ether | Not available | GC-MS | 2 | | | | | ● | | | | | | |
| Cinnamoylglycine | HMDB0011621 | LC-MS | 1[b] | | | ● | | | | | | | | |
| Methyluric acid | HMDB0003099 | LC-MS | 1[b] | | ● | | | | | | | | | |
| Dimethyluric acid | HMDB0001857 | LC-MS | 1[b] | | ● | | | | | | | | | |

[a]FB, fermented beverages; FCG, fermented cereals, and grains; FD, fermented dairy; GC−MS, gas chromatography−mass spectrometry; ID, identification; LC−MS, liquid chromatography−mass spectrometry; WG, whole grain; WT, white. Based on Spearman's correlations this metabolite is negatively associated with the fermented foods and food groups indicated but discriminant based on statistical significance in other tests (*e.g.*, PLS-DA and random forest). [b]Matched on retention time, mass, and fragmentation pattern.

various fermented foods and food groups, along with their platforms and biosamples of detection, are presented in Table 3 (plasma) and Table 4 (urine). The majority of the identified metabolites corresponded to the intakes of total FB (7 plasma and 19 urine), which encompasses coffee (3 plasma and 9 urine), wine (3 plasma and 10 urine), and beer (1 plasma and 6 urine). One urinary metabolite identified was discriminant for the intakes of total FCG, and one plasma metabolite was for white bread. However, metabolites discriminant for the intakes of whole-grain bread, cocoa, total FD, cheese, and yoghurt could not be identified.

A closer examination revealed that several of the metabolites (plasma dodecanoic acid, urinary D-psicose, glycine, D-gluconate, *m*-cresol, D-fucitol, and 2-keto-l-gluconic acid) were negatively associated with the fermented foods and food groups indicated (based on Spearman's correlations, Table S6) but contributed to the discrimination of the intake of these fermented foods based on statistical significance in



https://doi.org/10.1021/acs.jafc.2c05669
J. Agric. Food Chem. 2023, 71, 4426−4439

multivariate tests that do not distinguish between features that are at a higher abundance in a higher tertile intake group (*e.g.*, PLS-DA and random forest). Thus, these metabolites may not be suitable for reflecting fermented food intake and thus are not further discussed as FIBs. However, they may still be important biomarkers in revealing the metabolic effects of consuming these fermented foods.

Several of the identified FIBs overlapped across several fermented foods. Specifically, urinary 2,3-dihydroxybutanoic acid, ethyl α-D-glucopyranoside, and 4-hydroxybenzeacetic acid were discriminant for the intake of total FB, wine, and beer. Several other urinary metabolites also appeared to overlap between two fermented food groups, including 2,3-dihydroxypropyl phosphoric acid and D-lactose (total FB, beer), catechol, furoylglycine, niacin, and 3-deoxy-D-*ribo*-hexonic acid γ-lactone (total FB, coffee), as well as erythritol, tartaric acid, and arabinofuranose (total FB, wine). However, in each case, the significance of the total FB group could be driven by the significance of the individual beverages in this group. Similar overlaps were observed in plasma for L-cysteine (total FB, beer), xylitol (total FB, wine), and quinate (total FB, coffee). One metabolite identified for wine intake (erythritol) was identified in both plasma and urine.

**Stability of the Identified Biomarkers.** Spearman's correlations between the identified FIBs for fermented foods across different times between biosample collection and dietary assessment with the FFQ are presented in Table S6. Almost all correlations observed for the identification cohort (FFQ ± 14 d) remained significant with increasing time between biosample collection and FFQ completion (FFQ ± 30 d, 90 d, 180 d, and all FFQ), with only slight attenuations when the time between biosample collection and FFQ completion increased. The strongest correlations were observed between self-reported coffee intake and a series of FIBs, including plasma quinate, urinary niacin, furoylglycine, methyluric acid, and dimethyluric acid ($r_s \geq 0.4$, $p \leq 0.05$). For wine, the strongest correlations included urinary tartaric acid and arabinofuranose ($r_s \sim 0.4$), and for beer, the strongest correlation observed was ethyl α-D-glucopyranoside ($r_s \sim 0.27$) ($p \leq 0.05$). These correlations were also largely echoed between these metabolites and the intake of total FB. For total FCG, a significant moderate correlation was observed between self-reported intake and urinary glyceryl−glycoside ether in the identification cohort ($r_s \sim 0.37$), but the correlation attenuated in the full metabolomics cohort ($r_s < 0.3$) ($p \leq 0.05$). Conversely, correlations for intakes of cocoa, total FD, cheese, yoghurt, white bread, and whole-grain bread and their potential FIBs were either weak or nonexistent.

**Associations between Identified Biomarkers and Cardiometabolic Health Parameters.** The results of all associations between the identified biomarkers and CMD risk factors are presented in Table S7. In the fully adjusted model, 21 metabolites were positively associated and 11 were negatively associated with CMD risk parameters (unadjusted $p \leq 0.05$). After adjusting for multiple comparisons, 11 associations remained significant, including between plasma glutamic acid and urinary 2,3-dihydroxypropyl phosphoric acid with BMI (standardized (Std.) $\beta = 0.28$, $R^2 = 0.32$; Std. $\beta = 2.2 \times 10^{-7}$, $R^2 = 0.30$, respectively) and waist circumference (Std. $\beta = 1.73 \times 10^{-7}$, $R^2 = 0.50$; Std. $\beta = 0.28$, $R^2 = 0.48$, respectively) (FDR $p \leq 0.05$). Additional FDR-adjusted significant associations were observed between plasma xylitol (Std. $\beta = 2.20 \times 10^{-7}$, $R^2 = 0.26$), glutamic acid (Std. $\beta = 0.31$, $R^2 = 0.30$), and trigonelline (Std. $\beta = 0.34$, $R^2 = 0.28$), as well as urinary niacin (Std. $\beta = 2.55 \times 10^{-7}$, $R^2 = 0.32$), furoylglycine (Std. $\beta = 8.94 \times 10^{-8}$, $R^2 = 0.29$), and methyluric acid (Std. $\beta = 0.34$, $R^2 = 0.28$), with SCORE (FDR $p \leq 0.05$). A negative association was observed between plasma cinnamoylglycine with HbA1c (Std. $\beta = -0.27$, $R^2 = 0.36$, FDR $p \leq 0.05$). No FDR-adjusted significant associations were observed between metabolites with plasma lipids, glucose, or blood pressure.

## DISCUSSION

**FIBs Identified for the Habitual Intake of Individual Fermented Foods.** In the current work, we aimed to identify FIBs for fermented foods consumed in the habitual Dutch adult diet, which included coffee, wine, beer, whole-grain bread, white bread, cheese, yoghurt, and cocoa. A total of 12 plasma and 27 urinary metabolites were identified at level 1 or 2 from nontargeted GC−MS and LC−MS analyses, the majority of which corresponded to the intakes of coffee, wine, and beer (no metabolites were identified for cocoa, white bread, whole-grain bread, cheese, and yoghurt intake). These fermented foods were also coincidentally those with the highest intakes (in g/day) in the Dutch adult diet and span a wide range of intakes, which is conducive for the selection of discriminant metabolites. Several of the most promising FIBs identified for these foods were also previously captured by other nontargeted and targeted studies. For instance, plasma/serum quinate and trigonelline, as well as urinary niacin, furoylglycine, catechol, and methyluric and dimethyluric acids, have been previously reported as candidate FIBs of habitual coffee intake.[53−64] Out of the metabolites identified for wine intake, hydroxy(*iso*)butyric acid has been previously detected in serum after long-term (>4 weeks) wine intake,[65] tartaric acid in urine following acute wine intake,[66,67] and urinary 4-hydroxybenzeneacetic acid in urine following both acute and long-term (>4 weeks) wine intake.[65,68−71] The detection of these previously identified FIBs in our free-living population further supports their status as reliable indicators of the habitual intake of these fermented foods.

Additionally, several metabolites were identified for coffee, wine, and beer intake which have not been previously reported. For instance, we found urinary 3-deoxy-D-*ribo*-hexonic acid γ-lactone to be discriminant for coffee intake. This compound is a degradative product of glucose produced during the Maillard reaction,[72] which could have formed during coffee brewing. For wine intake, plasma xylitol, plasma/urinary erythritol, and urinary glucuronic acid, citramalate, 2,3-dihydroxybutanoic acid, arabinofuranose, and ethyl α-D-glucopyranoside were identified as potential FIBs. Furthermore, for beer intake, plasma L-cysteine, urinary 2,3-dihydroxypropyl phosphoric acid, 4-hydroxybenzeneacetic acid, D-lactose, 2,3-dihydroxybutanoic acid, and ethyl α-D-glucopyranoside were identified. While not detected previously in biofluids, almost all of these metabolites have been detected or quantified in the associated foods themselves. Erythritol (a natural sugar alcohol) has been previously detected in multiple fermented foods, including wine, beer, sake, coffee, cheese, and soy sauce.[54,73−75] Interestingly, erythritol can be produced by microorganisms (*e.g.*, *Penicillium* sp. used in the ripening of cheese).[15,75] Similarly, citramalate (a microbial metabolite that is found to be a byproduct of *Saccharomyces*, *Propionibacterium acnes*, and *Aspergillus niger*) has been detected in red wine.[76,77] The detection of these metabolites in the plasma and urine



https://doi.org/10.1021/acs.jafc.2c05669
J. Agric. Food Chem. 2023, 71, 4426−4439

metabolomes of free-living individuals consuming these fermented foods indicates that "fermentation-dependent" metabolites could act as powerful complementary FIBs (in addition to other FIBs originating from the raw food substrate) to improve the accuracy of dietary assessment of fermented foods in future studies. Thus, further validation studies are required in order to confirm the robustness and reliability of these newly identified FIBs (e.g., from the "identification cohort") in a separate population (e.g., "validation cohort").

One major challenge to the validation of single FIBs relates to their nonspecificity for a particular food. Indeed, the vast majority of metabolites identified for the intake of coffee, wine, and beer as described above have also been detected in biofluids following the consumption of other foods. For instance, plasma/serum quinic acid and urinary furoylglycine have also been identified for habitual cocoa intake,[78,79] while methyluric and dimethyluric acids, being caffeine metabolites, have naturally also been identified for the intake of caffeinated foods (*i.e.*, cocoa and tea).[79,80] The phenolic 4-hydroxybenzeneacetic acid corresponding to wine and beer intake has also been detected in urine after acute bread intake[81] and in serum after long-term coffee intake.[54] Furthermore, tartaric acid, a fairly specific FIB for wine intake, has also been identified in urine following acute and long-term bread intake[81] as well as acute beer intake.[82] The limitations of using these single metabolites as FIBs could be circumvented by developing reliable multimarker panels.[5] On the other hand, for fermented foods, nonspecific markers shared between different foods could also be useful for indicating common raw materials, fermentation processes (*e.g.*, lactic, acetic, alcoholic, or alkaline fermentation), and/or fermentation with common microorganisms (*e.g.*, lactic acid bacteria, with yeast). This work could be further extended by pathway analyses for the identified compounds, which could aid greatly in understanding their relationships to other metabolites and to human health.

**FIBs Identified for the Intake of Groups of Fermented Foods.** In the current work, we also explored using dry matter as a novel method to unify individual fermented foods with similar qualities into fermented food groups. Several groups were generated: total FB (comprising coffee, wine, and beer), total FCG (whole-grain and white bread), and total FD (cheese and yoghurt). By far, the largest number of identified metabolites were discriminant for the intake of total FB; however, the significance of the majority of these metabolites appeared to be largely driven by the individual beverages under this group. A few metabolites (plasma 2-hydroxybutyric acid, *trans*-aconitic acid, L-phenylalanine, and isoleucine; urinary guaiacol, 3-hydroxyhippuric acid, and 3,4-dihydroxyhydrocinnamic acid) appeared to be uniquely discriminant for total FBs. One metabolite was identified for total FCG intake (urinary glyceryl−glycoside ether). This metabolite has not been identified as a FIB previously and needs to be validated in further studies. No metabolites were identified for the intakes of cocoa or total FD, which could be due to the low or inconsistent intake of these foods (in the case of cocoa), or the discriminant metabolites being also of endogenous origin and thus influenced more heavily by human metabolism (in the case of fermented dairy).

While this is the first study to identify FIBs in the context of fermented food groups, this is also an area in need of further development. We formed the fermented food groups based on the groups for which the FFQ was previously validated for.[8,15] Evidently, there could be other strategies to group fermented foods, which could reveal different sets of FIBs. For instance, fermented foods could be grouped based on a common fermentation process (*e.g.*, lactic fermented foods and yeast-fermented foods), which may further reveal fermentation-dependent FIBs. Unfortunately, we did not have access to information on the fermenting microorganisms in order to group fermented foods according to this strategy. In addition, while we did not consider a total fermented food group in the current study, a total intake group might be worth exploring in future which would be highly relevant to examine the health impacts of a dietary pattern of fermented foods.

**Methodological Considerations for the Identification and Stability of the Biomarkers.** Although the primary aim of this study was to identify FIBs for the habitual intake of fermented foods, this work also contributes several methodological insights. First, to comprehensively capture the metabolome for FIB identification, we analyzed two biosamples (plasma and urine) using two analytical platforms (GC−MS and LC−MS). The 24 h urine samples were anticipated to better capture FIBs than plasma collected under fasting conditions, as depending on the speed of metabolism, the metabolite may not be detected even several hours after ingestion in plasma. Indeed, a larger number of urinary metabolites were significant and identified. Still, the number of significant urinary metabolites likely represented a smaller fraction of the total significant (and biologically relevant) metabolites in urine but which are present at relatively low concentrations due to dilution (a necessary step to ensure metabolites are measured within the linear range of the MS instruments). In addition, there could be differences in the metabolism of different metabolites that influence the detection of potential FIBs (*i.e.*, not all metabolites are eliminated in urine). A combination of these factors may explain why only one identified metabolite (erythritol for wine intake) overlapped between plasma and urine.

We also exploited a combination of univariate and multivariate statistical tests to identify the most discriminant FIBs—a strategy that has been explored by an increasing number of groups.[83−85] While the results of univariate and multivariate analyses are not always congruent, the use of both statistical approaches can generate complementary sets of FIBs. However, the results should be interpreted within the statistical framework from which they have been generated.[86] Additionally, we ran our analyses in positive mode to make use of the optimized settings by which the widest range of metabolites are expected to ionize and thus be detected. However, we recognize that running the same samples through negative mode (which was not possible due to time and economic constraints) could have been beneficial for expanding the detection of different metabolites and aid in metabolite identification.

Additionally, we investigated the stability of the identified FIBs with increasing time frames between biosample collection and completion of self-reported dietary assessment. This was a necessary analysis since the biosample collection did not occur at the same time as the dietary assessment and could be a source of variability. Importantly, we observed excellent stability (correlation coefficient and significance were maintained) with increasing time from biosample collection to the FFQ completion (within 14 d, 30 d, 180 d, and all) for almost all of the identified FIBs. The driving force for this stability could be the larger numbers of participants in longer time



https://doi.org/10.1021/acs.jafc.2c05669
J. Agric. Food Chem. 2023, 71, 4426−4439

frames (affording more statistical power). Moreover, these results could indicate that the FFQ used to collect information on self-reported fermented food intake in this study is fairly robust and/or that the diets of this population are very stable.

**Associations between Identified FIBs and CMD Risk Parameters.** We further examined associations between the identified FIBs with several CMD risk factors as a preliminary analysis to unravel the complex relationships between fermented food consumption and cardiometabolic health. Of the 39 metabolites identified, 7 were positively and 1 was negatively significantly associated with CMD risk factors after adjustment for multiple comparisons. All associations were weak, which may be attributed to the relatively healthy study population that may not have provided the gradient of CMD risk required to observe a large effect size. Thus, these associations need to be confirmed in larger, prospective cohorts or populations with a more distinctive divide between low and high CMD risk. Nonetheless, some of the associations we found have been reported in the literature. For instance, plasma glutamic acid has been positively associated with obesity, particularly with metabolically unhealthy obese phenotypes.[87] Other associations are more contested. In some studies, the consumption of non-nutritive sweeteners (which includes xylitol) has been shown to increase weight and waist circumference, as well as the incidence of obesity, hypertension, metabolic syndrome, type II diabetes, and cardiovascular events.[88] However, several recent systematic reviews have also revealed that the use of non-nutritive sweeteners (instead of sugar) reduces energy intake as well as body weight.[89,90] In all studies, the distinct effects of xylitol (compared to other non-nutritive sweeteners) as well as the underlying mechanisms behind these associations have yet to be verified. Similarly, while we observed significant positive associations between urinary niacin and plasma trigonelline with overall CMD risk (SCORE), this is not in line with the literature for specific CMD risk factors. Niacin is used as pharmacotherapy to prevent cardiovascular disease by lowering cholesterol levels in blood.[91] In mechanistic studies, trigonelline has been shown to improve insulin sensitivity by interfering with NADPH oxidase gene expression of pathways and mitochondrial electron chain transport.[92] Given these conflicting findings, further studies in larger, prospective cohorts are needed to clarify these associations and examine whether they play an intermediate role in CMD.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jafc.2c05669.

Classification of fermented foods from NQplus; summary of the significant metabolites for univariate and multivariate statistical analyses; list of suppliers of analytical standards; metabolites prioritized for identification from GC−MS and LC−MS; Spearman's correlations between identified metabolites and self-reported fermented food intakes (PDF)

Results of multiple linear regression (Table S7) (XLSX)

## ■ AUTHOR INFORMATION


### Corresponding Author

**Guy Vergères** − *Agroscope, CH-3003 Bern, Switzerland;* orcid.org/0000-0003-4574-0590; Phone: +41 (0)58 463 81 54; Email: guy.vergeres@agroscope.admin.ch

### Authors

**Katherine J. Li** − *Division of Human Nutrition and Health, Department of Agrotechnology and Food Science, Wageningen University & Research, 6700 AA Wageningen, The Netherlands; Agroscope, CH-3003 Bern, Switzerland*

**Kathryn J. Burton-Pimentel** − *Agroscope, CH-3003 Bern, Switzerland*

**Elske M. Brouwer-Brolsma** − *Division of Human Nutrition and Health, Department of Agrotechnology and Food Science, Wageningen University & Research, 6700 AA Wageningen, The Netherlands*

**Carola Blaser** − *Agroscope, CH-3003 Bern, Switzerland*

**René Badertscher** − *Agroscope, CH-3003 Bern, Switzerland;* orcid.org/0000-0001-9318-6470

**Grégory Pimentel** − *Agroscope, CH-3003 Bern, Switzerland*

**Reto Portmann** − *Agroscope, CH-3003 Bern, Switzerland*

**Edith J. M. Feskens** − *Division of Human Nutrition and Health, Department of Agrotechnology and Food Science, Wageningen University & Research, 6700 AA Wageningen, The Netherlands*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jafc.2c05669



### Funding

Funding to conduct the validation analyses was provided by Agroscope (Bern, Switzerland) for the Cardioferment Project. NDARD/NQplus was core funded by ZonMw (ZonMw, Grant 91110030). NDARD/NQPlus were also supported by Wageningen University (Wageningen, The Netherlands); add-on funding ZonMW Gezonde Voeding DHD-index (ZonMw, Grant 115100007); and add-on validation of BBMRI FFQ and Maastricht FFQ (grants BBMRI-NL RP9 and CP2011-38).


### Notes

The authors declare no competing financial interest.

## ■ ABBREVIATIONS

CMD, cardiometabolic disease; FB, fermented beverage; FCGs, fermented cereals and grains; FD, fermented dairy; FDR, false discovery rate; FFQ, food frequency questionnaire; FIB, food intake biomarker; MetS, metabolic syndrome; MSI, Metabolomics Standards Initiative; NDARD, National Dietary Assessment Reference Database; NQplus, Nutritional Questionnaire plus; PCA, principal component analysis; PLS-DA, partial least-square discriminant analysis; QC, quality control; SCORE, Systematic COronary Risk Evaluation; VIP, variable importance in projection

## ■ REFERENCES


(1) Bingham, S. A.; Gill, C.; Welch, A.; Day, K.; Cassidy, A.; Khaw, K. T.; Sneyd, M. J.; Key, T. J.; Roe, L.; Day, N. E. Comparison of dietary assessment methods in nutritional epidemiology: weighed records v. 24 h recalls, food-frequency questionnaires and estimated-diet records. *Br. J. Nutr.* **1994**, *72*, 619−643.

(2) Kipnis, V.; Midthune, D.; Freedman, L.; Bingham, S.; Day, N. E.; Riboli, E.; Ferrari, P.; Carroll, R. J. Bias in dietary-report instruments and its implications for nutritional epidemiology. *Public Health Nutr.* **2002**, *5*, 915−923.

(3) Gibbons, H.; Brennan, L. Metabolomics as a tool in the identification of dietary biomarkers. *Proc. Nutr. Soc.* **2017**, *76*, 42−53.
(4) Brouwer-Brolsma, E. M.; Brennan, L.; Drevon, C. A.; van Kranen, H.; Manach, C.; Dragsted, L. O.; Roche, H. M.; Andres-Lacueva, C.; Bakker, S. J. L.; Bouwman, J.; Capozzi, F.; De Saeger, S.; Gundersen, T. E.; Kolehmainen, M.; Kulling, S. E.; Landberg, R.; Linseisen, J.; Mattivi, F.; Mensink, R. P.; Scaccini, C.; Skurk, T.; Tetens, I.; Vergeres, G.; Wishart, D. S.; Scalbert, A.; Feskens, E. J. M. Combining traditional dietary assessment methods with novel metabolomics techniques: present efforts by the Food Biomarker Alliance. *Proc. Nutr. Soc.* **2017**, *76*, 619−627.
(5) Garcia-Aloy, M.; Rabassa, M.; Casas-Agustench, P.; Hidalgo-Liberona, N.; Llorach, R.; Andres-Lacueva, C. Novel strategies for improving dietary exposure assessment: Multiple-data fusion is a more accurate measure than the traditional single-biomarker approach. *Trends Food Sci. Technol.* **2017**, *69*, 220−229.
(6) Gibbons, H.; O'Gorman, A.; Brennan, L. Metabolomics as a tool in nutritional research. *Curr. Opin. Lipidol.* **2015**, *26*, 30−34.
(7) Borresen, E. C.; Henderson, A. J.; Kumar, A.; Weir, T. L.; Ryan, E. P. Fermented foods: patented approaches and formulations for nutritional supplementation and health promotion. *Recent Pat. Food, Nutr. Agric.* **2012**, *4*, 134−140.
(8) Li, K. J.; Brouwer-Brolsma, E. M.; Burton, K. J.; Vergères, G.; Feskens, E. J. M. Prevalence of fermented foods in the Dutch adult diet and validation of a food frequency questionnaire for estimating their intake in the NQplus cohort. *BMC Nutr.* **2020**, *6*, 69.
(9) Gille, D.; Schmid, A.; Walther, B.; Vergères, G. Fermented Food and Non-Communicable Chronic Diseases: A Review. *Nutrients* **2018**, *10*, 448.
(10) Mozaffarian, D. Dietary and Policy Priorities for Cardiovascular Disease, Diabetes, and Obesity. *Circulation* **2016**, *133*, 187−225.
(11) Tapsell, L. C. Fermented dairy food and CVD risk. *Br. J. Nutr.* **2015**, *113*, S131−S135.
(12) Mena-Sánchez, G.; Babio, N.; Martínez-González, M. A.; Corella, D.; Schröder, H.; Vioque, J.; Romaguera, D.; Martinez, J. A.; Lopez-Miranda, J.; Estruch, R.; Wärnberg, J.; Bueno-Cavanillas, A.; Serra-Majem, L.; Tur, J. A.; Arós, F.; Tinahones, F. J.; Sánchez, V. M.; Lapetra, J.; Pintó, X.; Vidal, J.; Vázquez, C.; Ordovás, J. M.; Delgado-Rodriguez, M.; Matía-Martín, P.; Basora, J.; Buil-Cosiales, P.; Fernandez-Carrion, R.; Fitó, M.; Salas-Salvadó, J.; PREDIMED-PLUS Investigators. Fermented dairy products, diet quality, and cardio-metabolic profile of a Mediterranean cohort at high cardiovascular risk. *Nutr., Metab. Cardiovasc. Dis.* **2018**, *28*, 1002−1011.
(13) Şanlier, N.; Gökcen, B. B.; Sezgin, A. C. Health benefits of fermented foods. *Crit. Rev. Food Sci. Nutr.* **2019**, *59*, 506−527.
(14) Lordan, R.; Tsoupras, A.; Mitra, B.; Zabetakis, I. Dairy Fats and Cardiovascular Disease: Do We Really Need to be Concerned? *Foods* **2018**, *7*, 29.
(15) Li, K. J.; Brouwer-Brolsma, E. M.; Burton-Pimentel, K. J.; Vergères, G.; Feskens, E. J. M. A systematic review to identify biomarkers of intake for fermented food products. *Genes Nutr.* **2021**, *16*, 5.
(16) Münger, L. H.; Garcia-Aloy, M.; Vazquez-Fresno, R.; Gille, D.; Rosana, A. R. R.; Passerini, A.; Soria-Florido, M. T.; Pimentel, G.; Sajed, T.; Wishart, D. S.; Andres Lacueva, C.; Vergères, G.; Pratico, G. Biomarker of food intake for assessing the consumption of dairy and egg products. *Genes Nutr.* **2018**, *13*, 26.
(17) Brouwer-Brolsma, E. M.; Streppel, M. T.; van Lee, L.; Geelen, A.; Sluik, D.; van de Wiel, A. M.; de Vries, J. H. M.; van 't Veer, P.; Feskens, E. J. M. A National Dietary Assessment Reference Database (NDARD) for the Dutch Population: Rationale behind the Design. *Nutrients* **2017**, *9*, 1136.
(18) Brouwer-Brolsma, E. M.; van Lee, L.; Streppel, M. T.; Sluik, D.; van de Wiel, A. M.; de Vries, J. H. M.; Geelen, A.; Feskens, E. J. M. Nutrition Questionnaires plus (NQplus) study, a prospective study on dietary determinants and cardiometabolic health in Dutch adults. *BMJ Open* **2018**, *8*, No. e020228.
(19) *The Dutch National Institute for Public Health and the Environment (RIVM)*; Nevo-Tabel. Nederlands Voedingsstoffenbestand, Voedingscentrum, Den Haag: The Netherlands, 2011. https://nevo-online.rivm.nl/.
(20) Willett, W. C. *Nutritional Epidemiology*, 3rd ed.; Oxford University Press: New York, New York, 2013.
(21) Allain, C. C.; Poon, L. S.; Chan, C. S.; Richmond, W.; Fu, P. C. Enzymatic determination of total serum cholesterol. *Clin. Chem.* **1974**, *20*, 470−475.
(22) Friedewald, W. T.; Levy, R. I.; Fredrickson, D. S. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin. Chem.* **1972**, *18*, 499−502.
(23) Mach, F.; Baigent, C.; Catapano, A. L.; Koskinas, K. C.; Casula, M.; Badimon, L.; Chapman, M. J.; De Backer, G. G.; Ference, V.; Graham, B. A.; Halliday, I. M.; Landmesser, A.; Mihaylova, U.; Pedersen, B.; Riccardi, T. R.; Richter, G.; Sabatine, D. J.; Taskinen, M. S.; Tokgozoglu, M. R.; Wiklund, L.; Wiklund, O. 2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. *Eur. Heart J.* **2020**, *41*, 111−188.
(24) Williams, B.; Mancia, G.; Spiering, W.; Agabiti Rosei, E.; Azizi, M.; Burnier, M.; Clement, D. L.; Coca, A.; de Simone, G.; Dominiczak, A.; Kahan, T.; Mahfoud, F.; Redon, J.; Ruilope, L.; Zanchetti, A.; Kerins, M.; Kjeldsen, S. E.; Kreutz, R.; Laurent, S.; Lip, G. Y. H.; McManus, R.; Narkiewicz, K.; Ruschitzka, F.; Schmieder, R. E.; Shlyakhto, E.; Tsioufis, C.; Aboyans, V.; Desormais, I. 2018 ESC/ESH Guidelines for the management of arterial hypertension. *Eur. Heart J.* **2018**, *39*, 3021−3104.
(25) Cosentino, F.; Grant, P. J.; Aboyans, V.; Bailey, C. J.; Ceriello, A.; Delgado, V.; Federici, M.; Filippatos, G.; Grobbee, D. E.; Hansen, T. B.; Huikuri, H. V.; Johansson, I.; Jüni, P.; Lettino, M.; Marx, N.; Mellbin, L. G.; Östgren, C. J.; Rocca, B.; Roffi, M.; Sattar, N.; Seferović, P. M.; Sousa-Uva, M.; Valensi, P.; Wheeler, D. C.; Piepoli, M. F.; Birkeland, K. I.; Adamopoulos, S.; Ajjan, R.; Avogaro, A.; Baigent, C.; Brodmann, M.; Bueno, H.; Ceconi, C.; Chioncel, O.; Coats, A.; Collet, J.-P.; Collins, P.; Cosyns, B.; Di Mario, C.; Fisher, M.; Fitzsimons, D.; Halvorsen, S.; Hansen, D.; Hoes, A.; Holt, R. I. G.; Home, P.; Katus, H. A.; Khunti, K.; Komajda, M.; Lambrinou, E.; Landmesser, U.; Lewis, B. S.; Linde, C.; Lorusso, R.; Mach, F.; Mueller, C.; Neumann, F.-J.; Persson, F.; Petersen, S. E.; Petronio, A. S.; Richter, D. J.; Rosano, G. M. C.; Rossing, P.; Rydén, L.; Shlyakhto, E.; Simpson, I. A.; Touyz, R. M.; Wijns, W.; Wilhelm, M.; Williams, B.; Aboyans, V.; Bailey, C. J.; Ceriello, A.; Delgado, V.; Federici, M.; Filippatos, G.; Grobbee, D. E.; Hansen, T. B.; Huikuri, H. V.; Johansson, I.; Jüni, P.; Lettino, M.; Marx, N.; Mellbin, L. G.; Östgren, C. J.; Rocca, B.; Roffi, M.; Sattar, N.; Seferović, P. M.; Sousa-Uva, M.; Valensi, P.; Wheeler, D. C.; Windecker, S.; Aboyans, V.; Baigent, C.; Collet, J.-P.; Dean, V.; Delgado, V.; Fitzsimons, D.; Gale, C. P.; Grobbee, D. E.; Halvorsen, S.; Hindricks, G.; Iung, B.; Jüni, P.; Katus, H. A.; Landmesser, U.; Leclercq, C.; Lettino, M.; Lewis, B. S.; Merkely, B.; Mueller, C.; Petersen, S. E.; Petronio, A. S.; Richter, D. J.; Roffi, M.; Shlyakhto, E.; Simpson, I. A.; Sousa-Uva, M.; Touyz, R. M.; Zelveian, P. H.; Scherr, D.; Jahangirov, T.; Lazareva, I.; Shivalkar, B.; Naser, N.; Gruev, I.; Milicic, D.; Petrou, P. M.; Linhart, A.; Hildebrandt, P.; Hasan-Ali, H.; Marandi, T.; Lehto, S.; Mansourati, J.; Kurashvili, R.; Siasos, G.; Lengyel, C.; Thrainsdottir, I. S.; Aronson, D.; Di Lenarda, A.; Raissova, A.; Ibrahimi, P.; Abilova, S.; Trusinskis, K.; Saade, G.; Benlamin, H.; Petrulioniene, Z.; Banu, C.; Magri, C. J.; David, L.; Boskovic, A.; Alami, M.; Liem, A. H.; Bosevski, M.; Svingen, G. F. T.; Janion, M.; Gavina, C.; Vinereanu, D.; Nedogoda, S.; Mancini, T.; Ilic, M. D.; Fabryova, L.; Fras, Z.; Jiménez-Navarro, M. F.; Norhammar, A.; Lehmann, R.; Mourali, M. S.; Ural, D.; Nesukay, E.; Chowdhury, T. A. 2019 ESC Guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD. *Eur. Heart J.* **2020**, *41*, 255−323.
(26) Alberti, K. G.; Eckel, R. H.; Grundy, S. M.; Zimmet, P. Z.; Cleeman, J. I.; Donato, K. A.; Fruchart, J. C.; James, W. P.; Loria, C. M.; Smith, S. C., Jr. Harmonizing the Metabolic Syndrome. *Circulation* **2009**, *120*, 1640−1645.

(27) Li, K. J.; Brouwer-Brolsma, E. M.; Fleuti, C.; Badertscher, R.; Vergères, G.; Feskens, E. J. M.; Burton-Pimentel, K. J. Associations between dairy fat intake, milk-derived free fatty acids, and cardiometabolic risk in Dutch adults. *Eur. J. Nutr.* **2023**, *62*, 185.
(28) Eisenmann, J. C. On the use of a continuous metabolic syndrome score in pediatric research. *Cardiovasc. Diabetol.* **2008**, *7*, 17.
(29) Jung, K. J.; Jee, Y. H.; Jee, S. H. Metabolic Risk Score and Vascular Mortality Among Korean Adults. *Asia Pac. J. Publ. Health* **2017**, *29*, 122−131.
(30) DeBoer, M. D.; Gurka, M. J. Clinical utility of metabolic syndrome severity scores: considerations for practitioners. *Diabetes, Metab. Syndr. Obes.: Targets Ther.* **2017**, *10*, 65−72.
(31) Conroy, R. M.; Pyorala, K.; Fitzgerald, A. P.; Sans, S.; Menotti, A.; De Backer, G.; De Bacquer, D.; Ducimetiere, P.; Jousilahti, P.; Keil, U.; Njolstad, I.; Oganov, R. G.; Thomsen, T.; Tunstall-Pedoe, H.; Tverdal, A.; Wedel, H.; Whincup, P.; Wilhelmsen, L.; Graham, I. M.; SCORE Project Group. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur. Heart J.* **2003**, *24*, 987−1003.
(32) Piepoli, M. F.; Piepoli, A. W.; Hoes, S.; Albus, C.; Brotons, C.; Catapano, A. L.; Cooney, M. T.; Corra, U.; Cosyns, B.; Deaton, C.; Graham, I.; Hall, M. S.; Hobbs, F. D.; Lochen, M. L.; Lollgen, H.; Marques-Vidal, P.; Perk, J.; Prescott, E.; Redon, J.; Richter, D. J.; Sattar, N.; Smulders, Y.; Tiberi, M.; van der Worp, H. B.; van Dis, I.; Verschuren, W. M.; Binno, S.; ESC Scientific Document Group. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts): Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). *Eur. J. Prev. Cardiol.* **2016**, *23*, NP1−NP96.
(33) Taylor, D. H., Jr.; Hasselblad, V.; Henley, S. J.; Thun, M. J.; Sloan, F. A. Benefits of smoking cessation for longevity. *Am. J. Public Health* **2002**, *92*, 990−996.
(34) Chinapaw, M. J.; Slootmaker, S. M.; Schuit, A. J.; van Zuidam, M.; van Mechelen, W. Reliability and validity of the Activity Questionnaire for Adults and Adolescents (AQuAA). *BMC Med. Res. Methodol.* **2009**, *9*, 58.
(35) Pimentel, G.; Burnand, D.; Münger, L. H.; Pralong, F. P.; Vionnet, N.; Portmann, R.; Vergères, G. Identification of Milk and Cheese Intake Biomarkers in Healthy Adults Reveals High Interindividual Variability of Lewis System-Related Oligosaccharides. *J. Nutr.* **2020**, *150*, 1058−1067.
(36) Li, K. J.; Burton-Pimentel, K. J.; Brouwer-Brolsma, E. M.; Feskens, E. J. M.; Blaser, C.; Badertscher, R.; Portmann, R.; Vergères, G. Evaluating the Robustness of Biomarkers of Dairy Food Intake in a Free-Living Population Using Single- and Multi-Marker Approaches. *Metabolites* **2021**, *11*, 395.
(37) Trimigno, A.; Münger, L.; Picone, G.; Freiburghaus, C.; Pimentel, G.; Vionnet, N.; Pralong, F.; Capozzi, F.; Badertscher, R.; Vergères, G. GC-MS Based Metabolomics and NMR Spectroscopy Investigation of Food Intake Biomarkers for Milk and Cheese in Serum of Healthy Humans. *Metabolites* **2018**, *8*, 26.
(38) Münger, L. H.; Trimigno, A.; Picone, G.; Freiburghaus, C.; Pimentel, G.; Burton, K. J.; Pralong, F. P.; Vionnet, N.; Capozzi, F.; Badertscher, R.; Vergères, G. Identification of Urinary Food Intake Biomarkers for Milk, Cheese, and Soy-Based Drink by Untargeted GC-MS and NMR in Healthy Humans. *J. Proteome Res.* **2017**, *16*, 3321−3335.
(39) 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 (HUSERMET) Consortium. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **2011**, *6*, 1060−1083.
(40) R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019. https://www.R-project.org/.
(41) Gromski, P. S.; Muhamadali, H.; Ellis, D. I.; Xu, Y.; Correa, E.; Turner, M. L.; Goodacre, R. A tutorial review: Metabolomics and partial least squares-discriminant analysis - a marriage of convenience or a shotgun wedding. *Anal. Chim. Acta* **2015**, *879*, 10−23.
(42) Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B* **1995**, *57*, 289−300.
(43) 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*, 13−19.
(44) Liaw, A.; Wiener, M. Classification and Regresson by randomForest. *R News* **2002**, *2*, 18−22.
(45) Altmann, A.; Toloşi, L.; Sander, O.; Lengauer, T. Permutation importance: a corrected feature importance measure. *Bioinformatics* **2010**, *26*, 1340−1347.
(46) Wright, M. N.; Ziegler, A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J. Stat. Software* **2017**, *77*, 1−17.
(47) Wishart, D. S.; Feunang, Y. D.; Marcu, A.; Guo, A. C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; Sayeeda, Z.; Lo, E.; Assempour, N.; Berjanskii, M.; Singhal, S.; Arndt, D.; Liang, Y.; Badran, H.; Grant, J.; Serra-Cayuela, A.; Liu, Y.; Mandal, R.; Neveu, V.; Pon, A.; Knox, C.; Wilson, M.; Manach, C.; Scalbert, A. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* **2018**, *46*, D608−D617.
(48) MassBank of North America (MoNA). Available at: http://mona.fiehnlab.ucdavis.edu.
(49) Smith, C. A.; Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* **2005**, *27*, 747−751.
(50) 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*, 1635−1638.
(51) 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.; Reily, M. D.; Thaden, J. J.; Viant, M. R. Proposed minimum reporting standards for chemical analysis. *Metabolomics* **2007**, *3*, 211−221.
(52) R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019. https://www.R-project.org/.
(53) Wang, Y.; Gapstur, S. M.; Carter, B. D.; Hartman, T. J.; Stevens, V. L.; Gaudet, M. M.; McCullough, M. L. Untargeted Metabolomics Identifies Novel Potential Biomarkers of Habitual Food Intake in a Cross-Sectional Study of Postmenopausal Women. *J. Nutr.* **2018**, *148*, 932−943.
(54) Cornelis, M. C.; Erlund, I.; Michelotti, G. A.; Herder, C.; Westerhuis, J. A.; Tuomilehto, J. Metabolomic response to coffee consumption: application to a three-stage clinical trial. *J. Intern. Med.* **2018**, *283*, 544−557.
(55) Guertin, K. A.; Loftfield, E.; Boca, S. M.; Sampson, J. N.; Moore, S. C.; Xiao, Q.; Huang, W. Y.; Xiong, X.; Freedman, N. D. A. J.; Cross, R.; Sinha, R. Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cancer. *Am. J. Clin. Nutr.* **2015**, *101*, 1000−1011.
(56) Heinzmann, S. S.; Holmes, E.; Kochhar, S.; Nicholson, J. K.; Schmitt-Kopplin, P. 2-Furoylglycine as a Candidate Biomarker of Coffee Consumption. *J. Agric. Food Chem.* **2015**, *63*, 8615−8621.

(57) Lang, R.; Wahl, A.; Stark, T.; Hofmann, T. Urinary N-methylpyridinium and trigonelline as candidate dietary biomarkers of coffee consumption. *Mol. Nutr. Food Res.* **2011**, *55*, 1613−1623.
(58) Rothwell, J. A.; Fillâtre, Y.; Martin, J. F.; Lyan, B.; Pujos-Guillot, E.; Fezeu, L.; Hercberg, S.; Comte, B.; Galan, P.; Touvier, M.; Manach, C. New biomarkers of coffee consumption identified by the non-targeted metabolomic profiling of cohort study subjects. *PLoS One* **2014**, *9*, No. e93474.
(59) Rothwell, J. A.; Keski-Rahkonen, P.; Robinot, N.; Assi, N.; Casagrande, C.; Jenab, M.; Ferrari, P.; Boutron-Ruault, M. C.; Mahamat-Saleh, Y.; Mancini, F. R.; Boeing, H.; Katzke, V.; Kühn, T.; Niforou, K.; Trichopoulou, A.; Valanou, E.; Krogh, V.; Mattiello, A.; Palli, D.; Sacerdote, C.; Tumino, R.; Scalbert, A. A Metabolomic Study of Biomarkers of Habitual Coffee Intake in Four European Countries. *Mol. Nutr. Food Res.* **2019**, *63*, No. e1900659.
(60) Chau, Y.-P.; Au, P. C. M.; Li, G. H. Y.; Sing, C. W.; Cheng, V. K. F.; Tan, K. C. B.; Kung, A. W. C.; Cheung, C. L. Serum Metabolome of Coffee Consumption and its Association With Bone Mineral Density: The Hong Kong Osteoporosis Study. *J. Clin. Endocrinol. Metab.* **2020**, *105*, No. e619.
(61) Hang, D.; Zeleznik, O. A.; He, X.; Guasch-Ferre, M.; Jiang, X.; Li, J.; Liang, L.; Eliassen, A. H.; Clish, C. B.; Chan, A. T.; Hu, Z.; Shen, H.; Wilson, K. M.; Mucci, L. A.; Sun, Q.; Hu, F. B.; Willett, W. C.; Giovannucci, E. L.; Song, M. Metabolomic Signatures of Long-term Coffee Consumption and Risk of Type 2 Diabetes in Women. *Diabetes Care* **2020**, *43*, 2588−2596.
(62) Shi, L.; Brunius, C.; Johansson, I.; Bergdahl, I. A.; Rolandsson, O.; Guelpen, B.; Winkvist, A.; Hanhineva, K.; Landberg, R. Plasma metabolite biomarkers of boiled and filtered coffee intake and their association with type 2 diabetes risk. *J. Intern. Med.* **2020**, *287*, 405−421.
(63) Playdon, M. C.; Sampson, J. N.; Cross, A. J.; Sinha, R.; Guertin, K. A.; Moy, K. A.; Rothman, N.; Irwin, M. L.; Mayne, S. T.; Stolzenberg-Solomon, R.; Moore, S. C. Comparing metabolite profiles of habitual diet in serum and urine. *Am. J. Clin. Nutr.* **2016**, *104*, 776−789.
(64) Midttun, O.; Ulvik, A.; Nygård, O.; Ueland, P. M. Performance of plasma trigonelline as a marker of coffee consumption in an epidemiologic setting. *Am. J. Clin. Nutr.* **2018**, *107*, 941−947.
(65) Vázquez-Fresno, R.; Llorach, R.; Alcaro, F.; Rodríguez, M. A.; Vinaixa, M.; Chiva-Blanch, G.; Estruch, R.; Correig, X.; Andrés-Lacueva, C. 1H-NMR-based metabolomic analysis of the effect of moderate wine consumption on subjects with cardiovascular risk factors. *Electrophoresis* **2012**, *33*, 2345−2354.
(66) Regueiro, J.; Vallverdú-Queralt, A.; Simal-Gándara, J.; Estruch, R.; Lamuela-Raventós, R. M. Urinary tartaric acid as a potential biomarker for the dietary assessment of moderate wine consumption: a randomised controlled trial. *Br. J. Nutr.* **2014**, *111*, 1680−1685.
(67) Regueiro, J.; Vallverdú-Queralt, A.; Simal-Gándara, J.; Estruch, R.; Lamuela-Raventós, R. Development of a LC-ESI-MS/MS approach for the rapid quantification of main wine organic acids in human urine. *J. Agric. Food Chem.* **2013**, *61*, 6763−6768.
(68) Vázquez-Fresno, R.; Llorach, R.; Urpi-Sarda, M.; Khymenets, O.; Bulló, M.; Corella, D.; Fitó, M.; Martínez-González, M.; Estruch, R.; Andres-Lacueva, C. An NMR metabolomics approach reveals a combined-biomarkers model in a wine interventional trial with validation in free-living individuals of the PREDIMED study. *Metabolomics* **2015**, *11*, 797−806.
(69) Vázquez-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 metabotypes after red wine polyphenol intake. *J. Nutr. Biochem.* **2016**, *28*, 114−120.
(70) Boto-Ordóñez, M.; Urpi-Sarda, M.; Queipo-Ortuño, M. I.; Corella, D.; Tinahones, F. J.; Estruch, R.; Andres-Lacueva, C. Microbial Metabolomic Fingerprinting in Urine after Regular Dealcoholized Red Wine Consumption in Humans. *J. Agric. Food Chem.* **2013**, *61*, 9166−9175.
(71) Pérez-Mañá, C.; Farré, M.; Rodríguez-Morató, J.; Papaseit, E.; Pujadas, M.; Fitó, M.; Robledo, P.; Covas, M. I.; Cheynier, V.; Meudec, E.; Escudier, J. L.; de la Torre, R. Moderate consumption of wine, through both its phenolic compounds and alcohol content, promotes hydroxytyrosol endogenous generation in humans. A randomized controlled trial. *Mol. Nutr. Food Res.* **2015**, *59*, 1213−1216.
(72) Haffenden, L. J.; Yaylayan, V. A. Nonvolatile oxidation products of glucose in Maillard model systems: formation of saccharinic and aldonic acids and their corresponding lactones. *J. Agric. Food Chem.* **2008**, *56*, 1638−1643.
(73) Shindou, T.; Sasaki, Y.; Miki, H.; Eguchi, T.; Hagiwara, K.; Ichikawa, T. Determination of erythritol in fermented foods by high performance liquid chromatography. *Sokuhin Eiseigaku Zasshi* **1988**, *29*, 419−422.
(74) Shindou, T.; Sasaki, Y.; Eguchi, H.; Euguchi, T.; Hagiwara, K.; Ichikawa, T. Identification of erythritol by HPLC and GC-MS and quantitative measurement in pulps of various fruits. *J. Agric. Food Chem.* **1989**, *37*, 1474−1476.
(75) Shindou, T.; Ishizuka, H. Quantitative determination of erythritol from various natural cheeses by HPLC. *Food Sci. Technol.* **1996**, *2*, 82−83.
(76) Hossain, A. H.; Hendrikx, A.; Punt, P. J. Identification of novel citramalate biosynthesis pathways in Aspergillus niger. *Fungal Biol. Biotechnol.* **2019**, *6*, 19.
(77) Carles, J. On the decarboxylations in wine and the appearance of citramalic acid. *Rev. Esp. Fisiol.* **1959**, *15*, 193−199.
(78) Pallister, T.; Jennings, A.; Mohney, R. P.; Yarand, D.; Mangino, M.; Cassidy, A.; MacGregor, A.; Spector, T. D.; Menni, C. Characterizing Blood Metabolomics Profiles Associated with Self-Reported Food Intakes in Female Twins. *PLoS One* **2016**, *11*, No. e0158568.
(79) Garcia-Aloy, M.; Llorach, R.; Urpi-Sarda, M.; Jáuregui, O.; Corella, D.; Ruiz-Canela, M.; Salas-Salvadó, J.; Fitó, M.; Ros, E.; Estruch, R.; Andres-Lacueva, C. A metabolomics-driven approach to predict cocoa product consumption by designing a multimetabolite biomarker model in free-living subjects from the PREDIMED study. *Mol. Nutr. Food Res.* **2015**, *59*, 212−220.
(80) Xie, G.; Zhao, A.; Zhao, L.; Chen, T.; Chen, H.; Qi, X.; Zheng, X.; Ni, Y.; Cheng, Y.; Lan, K.; Yao, C.; Qiu, M.; Jia, W. Metabolic fate of tea polyphenols in humans. *J. Proteome Res.* **2012**, *11*, 3449−3457.
(81) Zhu, Y.; Wang, P.; Sha, W.; Sang, S. Urinary Biomarkers of Whole Grain Wheat Intake Identified by Non-targeted and Targeted Metabolomics Approaches. *Sci. Rep.* **2016**, *6*, 36278.
(82) Monošík, R.; Dragsted, L. O. A versatile UHPLC-MSMS method for simultaneous quantification of various alcohol intake related compounds in human urine and blood. *Anal. Methods* **2016**, *8*, 6865−6871.
(83) Sachse, D.; Sletner, L.; Mørkrid, K.; Jenum, A. K.; Birkeland, K. I.; Rise, F.; Piehler, A. P.; Berg, J. P. Metabolic changes in urine during and after pregnancy in a large, multiethnic population-based cohort study of gestational diabetes. *PLoS One* **2012**, *7*, No. e52399.
(84) Szymańska, E.; Saccenti, E.; Smilde, A. K.; Westerhuis, J. A. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* **2012**, *8*, 3−16.
(85) Ellis, J. K.; Athersuch, T. J.; Thomas, L. D.; Teichert, F.; Pérez-Trujillo, M.; Svendsen, C.; Spurgeon, D. J.; Singh, R.; Järup, L.; Bundy, J. G.; Keun, H. C. Metabolic profiling detects early effects of environmental and lifestyle exposure to cadmium in a human population. *BMC Med.* **2012**, *10*, 61.
(86) Saccenti, E.; Hoefsloot, H. C. J.; Smilde, A. K.; Westerhuis, J. A.; Hendriks, M. M. W. B. Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* **2014**, *10*, 361−374.
(87) Bagheri, M.; Farzadfar, F.; Qi, L.; Yekaninejad, M. S.; Chamari, M.; Zeleznik, O. A.; Kalantar, Z.; Ebrahimi, Z.; Sheidaie, A.; Koletzko, B.; Uhl, O.; Djazayery, A. Obesity-Related Metabolomic Profiles and Discrimination of Metabolically Unhealthy Obesity. *J. Proteome Res.* **2018**, *17*, 1452−1462.

(88) Azad, M. B.; Abou-Setta, A. M.; Chauhan, B. F.; Rabbani, R.; Lys, J.; Copstein, L.; Mann, A.; Jeyaraman, M. M.; Reid, A. E.; Fiander, M.; MacKay, D. S.; McGavock, J.; Wicklow, B.; Zarychanski, R. Nonnutritive sweeteners and cardiometabolic health: a systematic review and meta-analysis of randomized controlled trials and prospective cohort studies. *CMAJ* **2017**, *189*, E929−E939.
(89) Normand, M.; Ritz, C.; Mela, D.; Raben, A. Low-energy sweeteners and body weight: a citation network analysis. *BMJ Nutr. Prev. Health* **2021**, *4*, 319−332.
(90) Rogers, P. J.; Hogenkamp, P. S.; de Graaf, C.; Higgs, S.; Lluch, A.; Ness, A. R.; Penfold, C.; Perry, R.; Putz, P.; Yeomans, M. R.; Mela, D. J. Does low-energy sweetener consumption affect energy intake and body weight? A systematic review, including meta-analyses, of the evidence from human and animal studies. *Int. J. Obes.* **2016**, *40*, 381−394.
(91) McKenney, J. New perspectives on the use of niacin in the treatment of lipid disorders. *Arch. Intern. Med.* **2004**, *164*, 697−705.
(92) Yoshinari, O.; Takenake, A.; Igarashi, K. Trigonelline ameliorates oxidative stress in type 2 diabetic Goto-Kakizaki rats. *J. Med. Food* **2013**, *16*, 34−41.





https://doi.org/10.1021/acs.jafc.2c05669
*J. Agric. Food Chem.* 2023, 71, 4426−4439