

Cite this: *Food Funct.*, 2024, 15, 1643

# Metabolome biomarkers linking dietary fibre intake with cardiometabolic effects: results from the Danish Diet, Cancer and Health-Next Generations MAX study†

Andrea Unión-Caballero,  ‡<sup>a,b</sup> Tomás Meroño, ‡<sup>a,b</sup> Raúl Zamora-Ros,<sup>c</sup> Agnetha Linn Rostgaard-Hansen,<sup>d</sup> Antonio Miñarro,<sup>b,e</sup> Alex Sánchez-Pla,<sup>b,e</sup> Núria Estanyol-Torres,<sup>a</sup> Miriam Martínez-Huelamo,<sup>a,b</sup> Marta Cubedo,  <sup>b,e</sup> Raúl González-Domínguez,<sup>a</sup> Anne Tjønneland,<sup>d</sup> Gabrielle Riccardi,<sup>f</sup> Rikard Landberg,<sup>g</sup> Jytte Halkjær\*<sup>d</sup> and Cristina Andrés-Lacueva\*<sup>a,b</sup>

Biomarkers associated with dietary fibre intake, as complements to traditional dietary assessment tools, may improve the understanding of its role in human health. Our aim was to discover metabolite biomarkers related to dietary fibre intake and investigate their association with cardiometabolic risk factors. We used data and samples from the Danish Diet Cancer and Health Next Generation (DCH-NG) MAX-study, a one-year observational study with evaluations at baseline, six and 12 months ( $n = 624$ , 55% female, mean age: 43 years, 1353 observations). Direct associations between fibre intake and plasma concentrations of 2,6-dihydroxybenzoic acid (2,6-DHBA) and indolepropionic acid were observed at the three time-points. Both metabolites showed an intraclass-correlation coefficient (ICC) > 0.50 and were associated with the self-reported intake of wholegrain cereals, and of fruits and vegetables, respectively. Other metabolites associated with dietary fibre intake were linolenoyl carnitine, 2-aminophenol, 3,4-DHBA, and proline betaine. Based on the metabolites associated with dietary fibre intake we calculated predicted values of fibre intake using a multivariate, machine-learning algorithm. Metabolomics-based predicted fibre, but not self-reported fibre values, showed negative associations with cardiometabolic risk factors (*i.e.* high sensitivity C-reactive protein, systolic and diastolic blood pressure, all FDR-adjusted  $p$ -values < 0.05). Furthermore, different correlations with gut microbiota composition were observed. In conclusion, 2,6-DHBA and indolepropionic acid in plasma may better link dietary fibre intake with its metabolic effects than self-reported values. These metabolites may represent a novel class of biomarkers reflecting both dietary exposure and host and/or gut microbiota characteristics providing a read-out that is differentially related to cardiometabolic risk.

Received 31st October 2023,  
Accepted 11th January 2024

DOI: 10.1039/d3fo04763f

rsc.li/food-function

## Introduction

Dietary fibre refers to plant-derived non-starch polysaccharides, resistant oligosaccharides, lignin, and resistant starch

that are resistant to human digestive enzymes.<sup>1</sup> National dietary guidelines include an optimal dietary fibre intake and recommendations for adults in Western countries are in the order of 30–35 g day<sup>-1</sup> for men and 25–32 g day<sup>-1</sup> for women.<sup>2</sup>

<sup>a</sup>Biomarkers and Nutrimetabolomics Laboratory, Department de Nutrició, Ciències de l'Alimentació i Gastronomia, Food Innovation Network (XIA), Institut de Recerca en Nutrició i Seguretat Alimentària (INSA-UB), Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona (UB), 08028 Barcelona, Spain.

E-mail: candres@ub.edu

<sup>b</sup>CIBER of Frailty and Healthy Aging (CIBERFES), Instituto de Salud Carlos III, 28029 Madrid, Spain

<sup>c</sup>Unit of Nutrition and Cancer, Cancer Epidemiology Research Program, Catalan Institute of Oncology (ICO), Bellvitge Biomedical Research Institute (IDIBELL), 08908 Barcelona, Spain

<sup>d</sup>Danish Cancer Society Research Center, Strandboulevarden 49, DK 2100 Copenhagen, Denmark

<sup>e</sup>Department of Genetics, Microbiology and Statistics, University of Barcelona, 08028 Barcelona, Spain

<sup>f</sup>Diabetes, Nutrition and Metabolism Unit, Department of Clinical Medicine and Surgery, Federico II University, 80138 Naples, Italy

<sup>g</sup>Department of Biology and Biological Engineering, Division of Food and Nutrition Science, Chalmers University of Technology, Gothenburg, Sweden

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d3fo04763f>

‡ Equally contributing authors.



However, overall average intakes are below recommendations in all countries.<sup>3</sup> Epidemiological studies have shown that dietary fibre intake is consistently associated with a reduced risk of weight gain, as well as for the incidence of non-communicable diseases such as cardiovascular diseases, type 2 diabetes, and some types of cancer.<sup>4</sup> Although the mechanisms for the beneficial effect exerted by dietary fibre intake are not fully understood, it could be partly explained by physiochemical properties (such as viscosity and bulking) affecting metabolic response to foods (such as glycemia, lipid profiles, *etc.*), and its ability to shape gut microbial composition and/or its metabolic activity.<sup>5–7</sup> Ultimately, this may cause the modulation of glycemia, blood lipids, and immune-metabolic pathways reducing chronic systemic inflammation,<sup>8–10</sup> a central hallmark in the development of non-communicable diseases.

In observational studies the accurate estimation of dietary fibre intake from dietary questionnaires is difficult and prone to systematic and random errors, principally due to the subjective nature of the assessment. Additionally, same food components are absorbed and metabolized differently in subjects with different gut microbiota or genetic backgrounds, making it difficult to accurately evaluate the correlations between the self-reported dietary fibre intake and the risk of chronic diseases.<sup>11</sup> Under such conditions, dietary biomarkers have emerged as a complementary strategy which could complement traditional dietary assessment and a framework for their anthology and validation have been developed.<sup>12</sup> Moreover, metabolite biomarkers may arise from the dietary fibre  $\times$  microbiota  $\times$  host interactions which may reflect and/or mediate differential risk profiles of non-communicable diseases.<sup>13,14</sup> Such biomarkers may be evaluated as novel biomarkers to guide future precision nutrition interventions aiming to maximize their health effects, through the identification of participants with particular host or gut microbiota metabolic characteristics that interact with dietary fibre exposure.<sup>15</sup>

To our knowledge, biomarkers associated with total dietary fibre are lacking. Nonetheless, some biomarkers of wholegrain intake have been suggested and evaluated since cereals comprise the main source of dietary fibre in most populations worldwide.<sup>3</sup> Among them, alkylresorcinols, a group of phenolic lipids present in the bran of wheat and rye, have been extensively assessed and applied as specific biomarkers of wholegrain wheat and rye intakes.<sup>16–18</sup> Both plasma alkylresorcinol concentrations as well as their metabolites in plasma and urine may reflect medium to long-term wholegrain wheat and rye intake in populations with a stable and frequent consumption.<sup>19</sup> Alkylresorcinol metabolites have longer half-lives than native alkylresorcinols, but their performance as biomarkers in free-living populations are similar.<sup>20,21</sup> Other biomarkers almost exclusively found in wheat and rye are benzoxazinoids and their phenylacetamide metabolites detected in urine and plasma after consumption of wholegrains.<sup>22</sup> These compounds are partly affected by food processing and gut microbial metabolism, thus hindering a direct association with the amount of wholegrain intake.<sup>23</sup> Avenanthramides and avenacosides have been suggested as specific biomarkers of oat intake,

but they appear to have short half-lives.<sup>24</sup> However, 2-amino-phenol sulfate, a gut microbial metabolite derived from benzoxazinoids, has been more extensively described to be elevated in both plasma and urine after high dietary fibre intake.<sup>21,25</sup> This biomarker has been scarcely tested over time and in free-living individuals in which is necessary before application in large epidemiologic studies.

Our main aim was to discover and validate metabolite biomarkers associated with dietary fibre intake and evaluate their performance or applicability to: (i) predict self-reported dietary fibre intake, and (ii) be associated with cardiometabolic risk factors in a validation sub cohort of the Danish Diet, Cancer and Health - Next Generations (DCH-NG),<sup>26</sup> the DCH-NG MAX study.

## Methods

### Study design and subjects

The DCH-NG cohort was established in Denmark between August 2015 and April 2019 and is an extension of the Diet, Cancer and Health (DCH) cohort.<sup>27</sup> The DCH-NG cohort includes 39 554 participants with complete data collection and involves biological children (Generation 1), their spouses (Generation 1-Parent), and the grandchildren (Generation 2) of the participants in DCH (Generation 0).<sup>28</sup> This analysis is based on a validation subsample called the DCH-NG MAX study, which consists of 720 participants aged 18 or older enrolled from August 2017 until January 2018. The main aims of DCH-NG MAX study are to validate a semi-quantitative food frequency questionnaire, analyze long-term reproducibility of plasma and urine metabolites, and assess the stability of the gut microbiota and its correlations with other measurements. Both questionnaire data and biological samples were collected at baseline, 6 and 12 months. All subjects completed two main questionnaires regarding lifestyle and dietary habits and participated in a health examination including the collection of biological samples, anthropometrics, and blood pressure measurements. Of the 720 participants enrolled in the DCH-NG MAX study at baseline, 676 had complete dietary data, and 624 had clinical measurements and biological samples available for metabolomics analyses. In total, the study comprised 1353 observations, out of which 380 were at 6 months, and 349 at 12 months (described with details in ESI Fig. 1†). The number of participants with complete clinical, dietary, and metabolomics data at baseline, 6 and 12 months was 287. The DCH-NG research project was approved by the Danish Data Protection Agency (journal number 2013-41-2043/2014-231-0094) and by the Committee on Health Research Ethics for the Capital Region of Denmark (journal number H-15001257). All participants provided their written informed consent before enrollment in the study.

### Anthropometric and blood pressure measurements

Anthropometric measurements were carried out with subjects wearing underwear and being barefoot. Height was measured



to the nearest 0.1 cm using a wireless stadiometer and weight was measured to the nearest 0.01 kg using a body composition analyzer (SECA mBCA515, Germany). Waist circumference was measured midway between the lower rib margin and iliac crest and to the nearest 0.1 cm. Visceral adipose tissue volume was measured with a DEXA-validated bioimpedance instrument (SECA mBCA515, Germany). Height, weight, and the bioelectrical impedance analysis were measured once, while the waist circumference was measured twice as a standard. If the difference between the two measurements for waist circumference was more than 1 cm, a third measurement was taken. Blood pressure and pulse rate were measured on the left arm, three times after at least 5 min of rest. The measurement with the lower systolic blood pressure was considered as valid.

### Dietary data

Participants in the DCH-NG MAX study completed a 24 h dietary recall (24-HDR) at each time point (baseline, 6 and 12 months) using the web-based tool myfood24 (<https://www.myfood24.org/>) developed by the Leeds University.<sup>29</sup> Data from myfood24 was linked primarily with the Danish National Food Database, which currently contains approximately 1600 Danish food items, including a recipe maker.<sup>30</sup> The participants reported all food consumed the day before the examination at the study center in grams by total portion size (as specified/selected by each participant). The portion sizes were based on reports from the Danish Food Institute.<sup>30</sup> The intake of macronutrients was expressed as percentage of calories using the energy equivalents for carbohydrates, proteins, and fats. Plant-based foods, as main sources of dietary fibre, were classified into 6 groups: total fruits (citrus fruits, dried fruits, preserved fruits, fruit juices, and other fruits), total vegetables (fruiting vegetables, leafy vegetables, cruciferous, stalk vegetables and sprouts, potatoes, other root vegetables, salads and prepared vegetable dishes, and vegetable juices), legumes, wholegrain cereals (porridge and ready-to-eat wholegrain cereals, wholegrain bread, wholegrain pasta, and wholegrain rice), refined cereals (breads, rice, pasta, bulgur, tortillas, ready-to-eat refined cereals, crackers, and cakes and biscuits), and nuts and seeds.

### Blood sampling and analysis of cardiometabolic risk factors

At baseline ( $n = 624$ ), six months ( $n = 380$ ), and twelve months ( $n = 349$ ), blood samples were drawn in tubes with lithium heparin at examination, and they were processed within 2 hours after extraction. Briefly, they were centrifuged for 10 min, 1800 g at 21 °C and stored overnight at 4 °C. The following day, they were separated and divided into aliquots of plasma and stored at -80 °C at the Danish National Biobank (DNB). For metabolomics analysis, plasma samples were shipped on dry ice to Chalmers University of Technology where they were stored at -80 °C prior to analysis.

Hemoglobin A1c (HbA1c), serum lipids and high sensitivity-C reactive protein (hsCRP) were measured in an auto-analyzer as previously described.<sup>31</sup>

### Metabolomics analysis of plasma samples

Plasma metabolomics analysis was performed following the semi-targeted procedure described by González-Domínguez *et al.*<sup>32</sup> In brief, plasma samples were prepared for UHPLC-MS/MS analysis by a protein precipitation protocol with ACN containing 1.5% v/v formic acid and 10 mM of ammonium formate. Sciex OS 2.1.6 software was used for data acquisition and processing.

### Metabolomics data preprocessing

Metabolomics data preprocessing was performed using the POMA R/Bioconductor package (<https://github.com/nutrimetabolomics/POMA>).<sup>33</sup> Data pre-processing included the removal of metabolites with more than 40% missing values, and those with a coefficient of variation (CV) > 30% in an internal quality control. The imputation of the remaining missing values was conducted using the KNN algorithm, the correction of batch-effects using the ComBat function ('sva' R package),<sup>34</sup> and data normalization using the auto-scaling. Afterwards, distances to the group centroid were computed based on Euclidean distances to remove outliers from the data matrix ( $\pm 1.5 \times \text{IQR}$ ). The working metabolomics dataset comprised 411 metabolites.

### Fecal microbiota

Stool samples were collected by the participants using a collection kit, including an EasySampler Stool Collector (GP Medical Devices, Denmark) and a sampling tube with a spoon-lid with no preservatives. Stool samples were stored at -80 °C prior to freeze-drying of the samples. Freeze-dried stool samples were used for extraction of DNA and 16S rRNA gene amplicon sequencing.

For DNA extraction, the QIAamp DNA Stool Mini kit from Qiagen GmbH (Hilden, Germany) was used. Bead beating was performed using 0.1 mm Silica/Zirconia beads from Bertin Technologies in Montigny-le-Bretonneux, France, and utilized a Precellys homogenizer from Bertin Technologies to conduct two rounds of bead-beating for 45 seconds. The V3 and V4 regions of the 16S rRNA gene were amplified with PCR using the 341F and 806R primers. An Illumina HiSeq 2500 Sequencer from Illumina in San Diego, CA, USA, was used for sequencing, producing 250 bp paired-end reads.

The deML pipeline was used to demultiplex 25 359 895 amplicon sequences and removal of primers, ambiguous sequences, and chimeric sequences using dada2 within QIIME2. The obtained amplicon sequence variants (ASVs) were assigned taxonomy using the Silva 16S rRNA Database at  $\geq 97\%$  identity. ASVs sharing taxonomic identity were merged, and phylotypes represented by a single read were excluded. Relative abundances of taxa were calculated as the proportion of reads for the taxon in percent of all reads for the individual, and the prevalence as the proportion (%) of individuals where the taxon was identified. For analyses we selected the genera with a relative abundance greater than 1% in the whole sample and with an intraclass correlation coefficient (ICC) >



0.50 among the three time-points of the study. The final dataset comprised 150 gut bacteria genera.

### Statistical analyses

Mean and standard deviation, or median (Q1–Q3) were used to describe variables following a Gaussian or skewed distribution, respectively. Number of participants (%) were used for categorical variables. Dietary fibre intake was categorized by tertiles at each evaluation and sociodemographic characteristics and cardiometabolic risk factors were compared using linear mixed models with random intercepts for the participant ID and adjusted for age, sex, BMI, and time of the evaluation (baseline, 6, or 12 months).

The associations between dietary fibre intake and metabolites were assessed by linear mixed models with random intercepts for the participant ID and adjusted for age, sex, BMI, and time of the evaluation. *P*-Values were adjusted for multiple comparisons using the Benjamini-Hochberg false discovery rate (FDR). An FDR-adjusted *p*-value <0.05 was considered significant. ICCs were estimated for dietary fibre intake and for selected metabolites using random intercept linear mixed models.<sup>35</sup>

Multivariate analyses were conducted using Mixed Graphical Models (MGM) with the 'mgm' R-package. MGMs are undirected probabilistic graphical models, where each node corresponds to one variable, and the edges between two nodes represent a conditional dependency between them given all other variables in the model.<sup>36</sup> MGM specifications were set to allow the maximum number of interactions in the network. Variables in the model were dietary fibre intake, age, sex, BMI and the whole metabolomic set of variables. Separate MGM models were fitted for the baseline, 6 months, and 12 months data. MUVR-PLS analysis<sup>37</sup> was conducted with the 'MUVR' R-package.

Associations between metabolomics-derived biomarkers and intakes of major and minor food groups were assessed using age-, sex- and BMI-adjusted linear mixed models with random intercepts and slopes (time). Similar analyses were conducted for the association between metabolomics-based predicted and self-reported fibre intake and cardiometabolic risk factors. Spearman correlations were conducted between self-reported fibre, metabolomics-based predicted fibre, and gut microbiota genera.

All statistical analyses were performed using R version 4.2.3 (R foundation, Austria).

## Results

### Study population and characteristics according to dietary fibre intake

Clinical, sociodemographic and cardiometabolic risk factors of the population according to self-reported dietary fibre intake tertiles during the one-year study are shown in Table 1. No statistically significant differences in sociodemographic and clinical characteristics were observed across tertiles of self-reported

dietary fibre intake, except for high-sensitivity C-reactive protein (hsCRP), where participants in the highest tertile were those with the lowest values (Table 1). Dietary characteristics are shown in the ESI Table 1.† Participants in the highest tertile of dietary fibre intake presented higher energy, polyunsaturated fatty acids (PUFAs), total carbohydrates (as % of energy) and sodium intake than the other participants. Moreover, they showed a lower consumption of saturated fatty acids (SFAs) and protein (as % of energy). Particularly, food groups with the highest contribution to total dietary fibre intake were wholegrain cereals and vegetables (41% and 22%, respectively), followed by fruits, refined cereal foods (including cakes and biscuits), nuts and seeds, and legumes (18%, 12%, 5%, and 1%, respectively). During the one-year study period, there were no statistically significant changes in dietary fibre intake (*p* < 0.05), and participants reported a stable intake of dietary fibre (ICC = 0.39, ESI Fig. 2†).

### Metabolites associated with dietary fibre intake

Association between dietary fibre intake and plasma metabolites are shown in Fig. 1. The following wholegrain-related metabolites were positively associated with total dietary fibre intake: 2,6-dihydroxybenzoic acid (2,6-DHBA); 3,4-dihydroxybenzoic acid (3,4-DHBA);<sup>38</sup> 3-(2-hydroxyphenyl)-propionic acid (2-HPPA); the benzoxazinoid metabolite, 2-aminophenol;<sup>39</sup> and pipercolic acid betaine.<sup>40,41</sup> In addition, the following metabolites related with plant foods were positively associated with total dietary fibre intake: conjugated enterolactones (both glucuronide and sulfate), that are lignans produced by gut microbial fermentation of dietary plant foods;<sup>42</sup> hippuric acid;<sup>43</sup> 4-hydroxyhippuric acid;<sup>44</sup> catechol-sulfate; proline betaine;<sup>45</sup> hypaphorine;<sup>46</sup> and  $\alpha$ -tocopherol.<sup>47</sup> Other metabolites positively associated with total dietary fibre intake included the gut microbial metabolite of tryptophan, indolepropionic acid;<sup>48</sup> a metabolite commonly related with olives and olive oil, hydroxytyrosol;<sup>49</sup> the amino acid aspartic acid; and linoleoyl-carnitine. Last, one metabolite was negatively associated with total dietary fibre, ethyl glucuronide, a metabolite associated with alcohol intake.<sup>50</sup>

When applying MGMs at baseline data, 4 metabolites (indolepropionic acid, 2,6-DHBA, 3,4-DHBA and linoleoyl-carnitine) were found associated with total dietary fibre intake after adjustment for age, sex, BMI, and the set of metabolites (Fig. 2). At 6 months, 2,6-DHBA, linoleoyl-carnitine and ethylglucuronide were associated with total dietary fibre intake, while at 12 months 2,6-DHBA and indolepropionic acid were significantly associated with total dietary fibre intake (Fig. 2). Similar results were obtained by other multivariate analysis using MUVR algorithm (ESI Fig. 3†). Indeed, 2,6-DHBA, linoleoyl-carnitine and indolepropionic acid were the metabolites with the highest importance in MUVR analysis.

Plots showing the stability of repeated measurements over time of these metabolites are shown in ESI Fig. 2.† In particular, 2,6-DHBA, and indolepropionic acid had an ICC > 0.50, which is considered moderately reliable and stable on repeated measurements over time.<sup>51</sup> On contrary, ICC of linolenoyl-car-



**Table 1** Clinical, sociodemographic and cardiometabolic risk factors of the DCH-NG MAX study population according to tertiles of dietary fibre intake

	All <i>n</i> = 624 <i>k</i> = 1353	Fibre <16 g day <sup>-1</sup> (T1) <i>k</i> = 451	Fibre: 16–25 g day <sup>-1</sup> (T2) <i>k</i> = 452	Fibre >25 g day <sup>-1</sup> (T3) <i>k</i> = 450
<b>Sociodemographic characteristics</b>				
Age (y)	43 ± 12	44 ± 12	43 ± 13	43 ± 13
Sex, female ( <i>n</i> , %)	742 (55)	251 (56)	275 (61)	216 (48)
BMI (kg m <sup>-2</sup> )	25 ± 4	26 ± 4	25 ± 4	24 ± 3
WC (cm)	88 ± 12	90 ± 12	87 ± 12	86 ± 12
VAT (l)	1.4 (0.7–2.4)	1.5 (0.9–2.8)	1.3 (0.7–2.2)	1.1 (0.7–2.3)
Physical activity ( <i>n</i> , %)				
Not regular	213 (16)	91 (20)	73 (16)	49 (11)
1/month last 6 months	104 (8)	37 (8)	37 (8)	30 (7)
1/month last 12 months	1035 (76)	322 (72)	342 (76)	371 (82)
Smoking status ( <i>n</i> , %)				
Never	732 (54)	238 (53)	233 (51)	261 (58)
Former	374 (28)	119 (26)	132 (29)	123 (27)
Current	246 (18)	93 (21)	87 (19)	66 (15)
<b>Cardiometabolic risk factors</b>				
DBP (mmHg)	81 ± 11	81 ± 11	80 ± 10	79 ± 11
SBP (mmHg)	117 ± 16	117 ± 16	115 ± 14	116 ± 16
HbA1c (mmol mol <sup>-1</sup> )	34.5 ± 6.0	35.2 ± 7.0	33.9 ± 5.0	34.2 ± 5.6
TG (mmol L <sup>-1</sup> )	1.1 (0.8–1.7)	1.1 (0.8–1.7)	1.0 (0.8–1.4)	1.1 (0.7–1.6)
TC (mmol L <sup>-1</sup> )	5.0 ± 1.0	5.1 ± 1.0	4.9 ± 1.0	4.9 ± 0.9
HDL-C (mmol L <sup>-1</sup> )	1.54 ± 0.42	1.54 ± 0.54	1.58 ± 0.43	1.56 ± 0.43
LDL-C (mmol L <sup>-1</sup> )	3.0 ± 0.9	3.1 ± 0.9	3.0 ± 0.9	2.9 ± 0.8
hsCRP (mg L <sup>-1</sup> )	0.7 (0.3–1.5)	0.9 (0.4–1.7)	0.8 (0.3–1.7)	0.5 (0.2–1.3) **

BMI, body mass index; WC, waist circumference; VAT, visceral adipose tissue; DBP, diastolic blood pressure; SBP, systolic blood pressure; HbA1c, hemoglobin A1c; TG, triglycerides; TC, total cholesterol; HDL-C, high-density lipoprotein-cholesterol; LDL-C, low-density lipoprotein cholesterol; hsCRP, high-sensitivity C-reactive protein. Mean and standard deviation, or median (Q1–Q3) were used to describe continuous variables following a Gaussian or skewed distribution, respectively. \**p* for trend <0.05, \*\**p* for trend <0.01, \*\*\**p* for trend <0.001 using generalized linear models adjusting for age-, sex- and BMI. Variables with skewed distribution were log-transformed before entering the analyses. *n* = number of subjects, *k* = total number of observations.

nitine was 0.08, showing high variability between measurements. Additionally, we analyzed the association between dietary sources of dietary fibre intake and metabolites selected in univariate and multivariate analyses (ESI Fig. 4†). Consumption of wholegrain cereals was directly associated with 2,6-DHBA, 2-aminophenol and linoleoyl-carnitine levels, while it was inversely associated with ethyl-glucuronide levels. Total vegetable intake was directly related with indolepropionic acid. Consumption of fruiting vegetables, leafy vegetables, and root vegetables other than potatoes were associated with 3,4-DHBA, and linoleoyl-carnitine. Total fruits intake was directly associated with proline betaine, indolepropionic acid, and 3,4-DHBA. Consumption of citrus fruit and fruit juices were positively related with proline betaine. Last, legumes intake was positively associated with 3,4-DHBA, and linoleoyl-carnitine (ESI Fig. 4†).

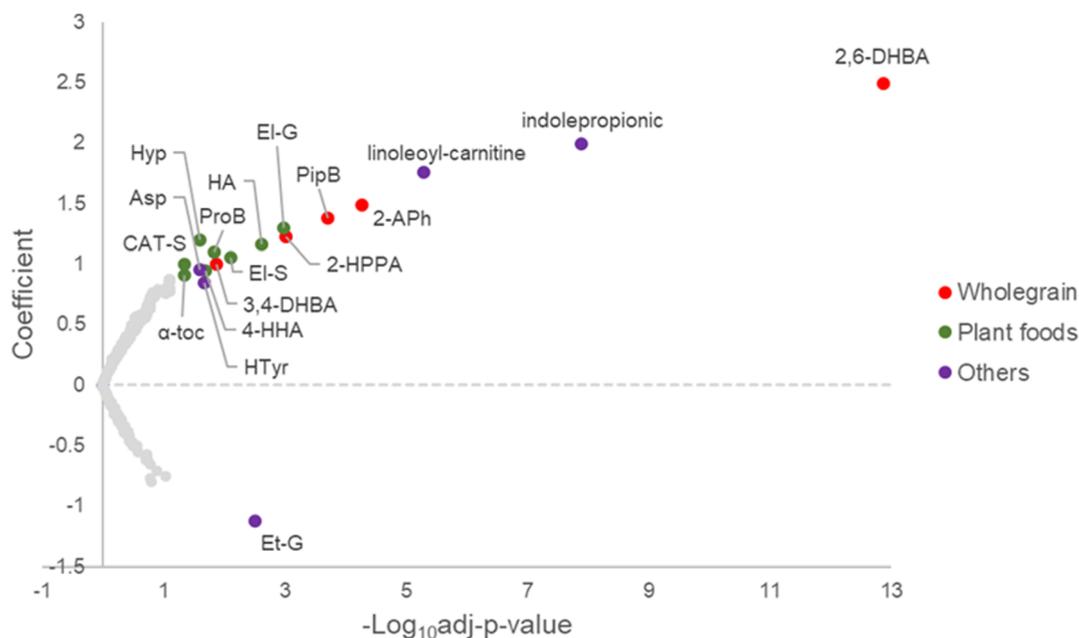
We estimated fibre intake values using the MGM models at each time-point and we compared the associations of metabolomics-based predicted and of self-reported fibre intake values with cardiometabolic risk factors (Fig. 3). Metabolomics-based predicted fibre intake was negatively associated with diastolic blood pressure, systolic blood pressure and high sensitivity-C reactive protein (hsCRP) (Fig. 3). Conversely, self-reported fibre intake showed no significant association with any cardiometabolic risk factor (Fig. 3).

To gain insight into the differences between self-reported and metabolomics-based predicted fibre, we analyzed their correlations with gut microbiota composition at the genus level. Fig. 4 shows the statistically significant correlations (FDR-adjusted *p*-value <0.05) for self-reported and metabolomics-based predicted fibre with gut microbiota genera. Although correlations with several genera were shared among both fibre variables, *Ruminococcaceae* (*UCG-002*, *UCG-005*, *UCG-013*), and *Eubacterium eligens* were positively correlated only with metabolomics-based predicted fibre. but not with the self-reported one. Interestingly, *Bacteroides*, a genus involved in polysaccharide metabolism,<sup>52</sup> was negatively correlated both with both self-reported and metabolomics-based predicted fibre.

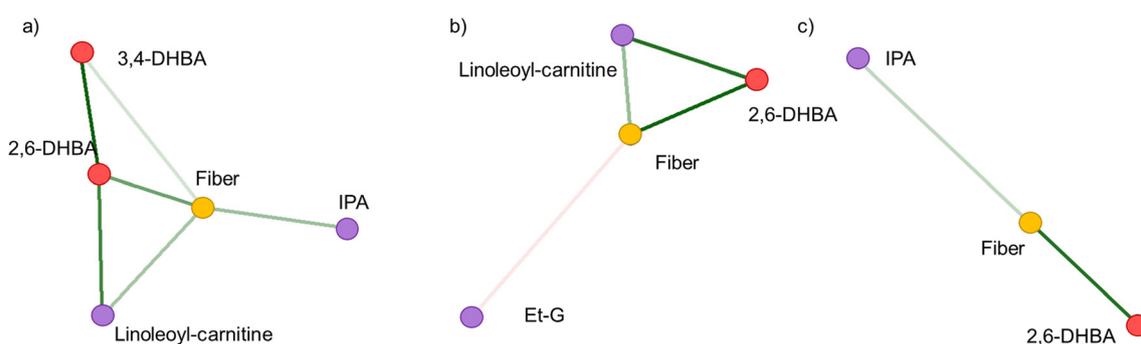
## Discussion

The present study positions indolepropionic acid and 2,6-DHBA as metabolite biomarkers linking dietary fibre intake with its cardiometabolic effects in an observational study with three repeated measurements over one year. For the first time, we report associations between self-reported dietary fibre intake and metabolites that were consistent for one year. Furthermore, we showed that a set of these metabolites may





**Fig. 1** Association between dietary fibre intake and plasma metabolites in the DCH-NG MAX study ( $n = 624$ ,  $k = 1353$ ). According to linear mixed models with random intercepts (defined by participant ID), adjusted for age, sex, BMI, and time. FDR-adjusted  $p$ -value  $< 0.05$  were considered significant. 2,6-DHBA, 2,6-dihydroxybenzoic acid; 2-APh, 2-aminophenol; PipB, pipercolic acid betaine, EI-G, enterolactone-glucuronide; 2-HPPA, 3-(2-hydroxyphenyl)-propionic acid; HA, hippuric acid; EI-S, enterolactone-sulfate; 3,4-DHBA, 3,4-dihydroxybenzoic acid; ProB, proline betaine; Hyp, hypaphorine; Asp, aspartic acid; HTyr, hydroxytyrosol; 4-HHA, 4-hydroxyhippuric acid; CAT-S, catechol-sulfate;  $\alpha$ -toc,  $\alpha$ -tocopherol; Et-G, ethyl glucuronide.  $n =$  number of subjects,  $k =$  total number of observations. Dot colours were assigned according to previous studies showing associations between metabolites and food groups.



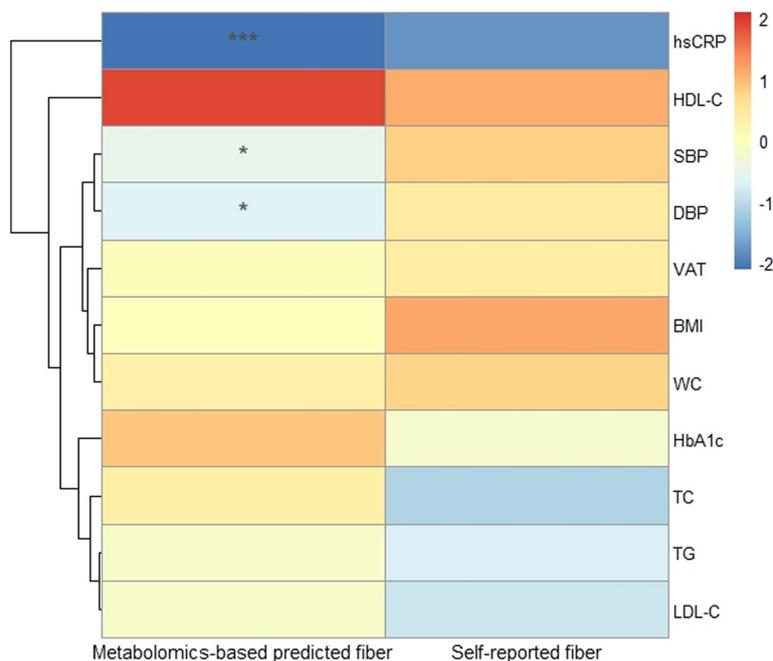
**Fig. 2** First-order neighborhood of dietary fibre intake according to Mixed Graphical Models in participants with metabolomic analyses at baseline (panel a), and 6 (panel b) and 12 months (panel c,  $n = 624$ ,  $k = 1353$ ). Edge intensity reflects the strength of an association from strong positive (dark green) to strong negative association (dark red). Variables included in the mixed graphical model were dietary fibre intake, age, sex, BMI, and all the metabolomic variables. 2,6-DHBA, 2,6-dihydroxybenzoic acid; 3,4-DHBA, 3,4-dihydroxybenzoic acid; IPA, indolepropionic acid; Et-G, ethyl-glucuronide.  $n =$  number of subjects,  $k =$  total number of observations. Dot colours were assigned according to previous studies showing associations between metabolites and food groups.

summarize information from diet, host and gut microbiota metabolism allowing the identification of individuals with altered levels of cardiometabolic risk factors. As such, we proposed these metabolites as part of a new group of biomarkers that reflect both dietary fibre exposure and, at least to a higher extent than self-reported fibre intake values, its metabolic effects.

The scientific rationale for dietary fibre intake recommendations relies on its health benefits.<sup>53</sup> However, the key physio-

logical effects elicited by dietary fibre are still not fully unrevealed, and the effects will vary depending on chemical and functional heterogeneity between different types of dietary fibre. In general, dietary fibre is known to contribute to shape gut microbiota composition and functionality, and to increase the output of fermentative end-products, such as short-chain fatty acids.<sup>54</sup> Thus, the beneficial effect of the interactions between dietary fibre and gut microbiota can result from the modulation of the production of microbial metabolites, and/or





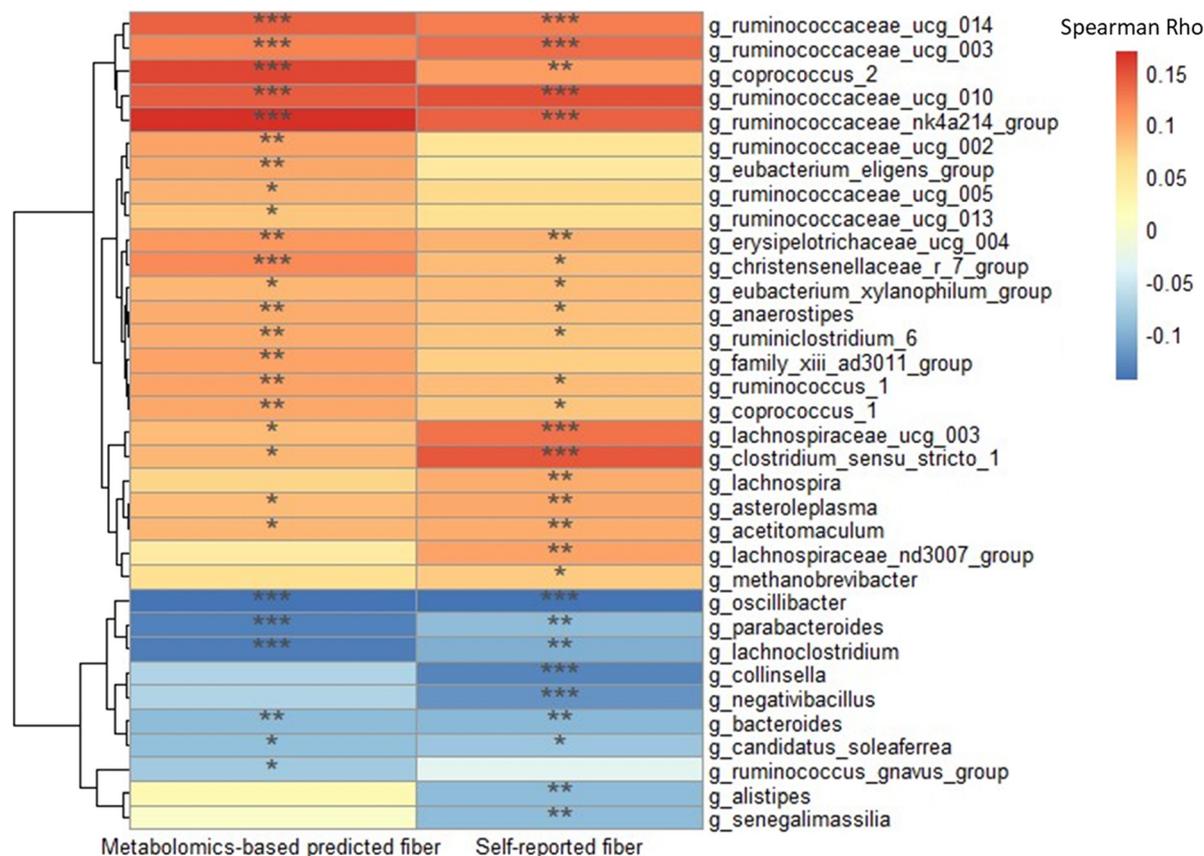
**Fig. 3** Heatmap showing the association between metabolomics-based predicted fibre intake and self-reported fibre intake with cardiometabolic risk factors in the DCH-NG MAX study. Coefficients calculated using age-, sex-adjusted linear mixed models with random intercepts and slopes (time) in the whole study population ( $n = 624$ ,  $k = 1353$ ). \*FDR-adjusted  $p$ -value  $< 0.05$ , \*\*\* FDR-adjusted  $p$ -value  $< 0.001$ . hsCRP, high sensitivity-C reactive protein; HDL-C, HDL-cholesterol; SBP, systolic blood pressure; DBP, diastolic blood pressure; VAT, visceral adipose tissue; BMI, body mass index; WC, waist circumference; HbA1c, hemoglobin A1c; TC, total cholesterol; TG, triglycerides; LDL-C, LDL-cholesterol.

from changes in the composition and functions of the gut microbiota supposed to mediate its health benefits.<sup>54</sup> Therefore, it is crucial to identify relevant biomarkers that could reflect the physiological effects of specific dietary fibre in combination with gut microbiota and host characteristics relevant to cardiometabolic health. In our analysis, we used a machine-learning algorithm (MGM) for selecting metabolites associated with dietary fibre, and these results were in line with those obtained by univariate analysis. The importance of these metabolites was confirmed using other machine-learning method for variable selection, such as MUV algorithm.<sup>37</sup> Furthermore, we used MGM to calculate predicted dietary fibre intake levels and, up to some extent, these values showed more relevant associations with cardiometabolic risk factors than self-reported dietary fibre intake values *per se*. In part, this could be a result from inaccurate self-reported dietary fibre intake. However, a physiological explanation can also be proposed in relation with different capacities of gut microbiota to metabolize differently complex carbohydrates, driving different effects in the host cardiometabolic health. In addition, we assessed gut microbiota correlations with both self-reported and metabolomics-based predicted dietary fiber intakes. We observed how both variables shared most of the significant correlations with gut microbiota at genus level. However, some genus such as *Ruminococcaceae* UCG-002, UCG-005, UCG-013, and *Eubacterium eligens*, were exclusively correlated with metabolomics-based predicted fibre. *Ruminococcaceae* UCG-005 has been specifically associated with

circulating levels of the short-chain fatty acid acetate,<sup>55</sup> and has also been reported to attenuate obesity.<sup>56</sup> In addition, its relative abundance has been described to increase in response to resistant starch intake in normal weight individuals.<sup>57</sup> *Eubacterium eligens* has been linked to a higher adherence to the Mediterranean diet<sup>58</sup> and it was suggested to exhibit a positive impact on intestinal inflammation.<sup>59</sup> Conversely, *Bacteroides*, one of the main polysaccharide degrading bacteria genus in humans,<sup>60</sup> negatively correlated with self-reported and predicted dietary fibre values. Several species of *Bacteroides* are well-known dietary fibre fermenters, but not having information at the species level may have been an important limitation at this point. Overall, these differences between self-reported and metabolomics-based fibre intake values may be relevant to explain divergent health effects and interindividual variability, since low- to high-inflammatory responses were observed during a high-fibre intervention.<sup>61</sup>

This latter report and our results suggest that some subjects may benefit (or not) from elevated intakes of dietary fibre due to an specific interaction between dietary fibre and host/gut microbiota characteristics. Future studies are needed to find metabolite profiles uniquely reflecting the interactions between these features and their relation to risk factors for diseases. Such profiles could be used to guide tailored fibre rich foods to those that benefit the most, *i.e.*, personalized nutrition. In summary, plasma levels of 2,6-DHBA and the microbial metabolite indolepropionic acid may represent a new type of surrogate biomarker that reflect the outcome of an





**Fig. 4** Heatmap showing the significant correlations between gut microbiota (genus level) and metabolomics-based predicted fibre and self-reported fibre ( $n = 561$ ,  $k = 1213$ ). \*FDR-adjusted  $p$ -value  $< 0.05$ , \*\* FDR-adjusted  $p$ -value  $< 0.01$ , \*\*\* FDR-adjusted  $p$ -value  $< 0.001$ . Only significant correlations (FDR-adjusted  $p$ -value  $< 0.05$ ) are plotted.

interaction between intestinal microbial community and fibre-rich plant-based foods that is beneficially associated with cardiometabolic risk markers. Further analyses of the gut microbiota characteristics using shotgun metagenomics are warranted. The metabolite and clinical biomarker profile need validation in other studies from different populations but holds the potential to be used for tailoring personalized dietary fibre interventions for cardiometabolic diseases prevention.

Discussion on the association between the individual metabolites and dietary fibre and previous studies with similar and supporting findings are in the ESI.† One of the strengths of the study is the use of a high-throughput, comprehensive targeted metabolomics method with three assessments during a one-year study period. For the first time we were able to report reliability indicators (reproducibility) of plasma microbial metabolites for one year. Moreover, our targeted metabolomics method showed a low coefficient of variation in quality control samples for these metabolites (5.1% for 2,6-DHBA and 6.2% for indolepropionic acid). On the other hand, the use of 24 h dietary recalls could be considered as an advantage because the quantification of intakes is more precise compared with food frequency questionnaires. However, a one day

24 h dietary recall may not reflect typical diet and are limited in its generalizability. Dietary 24 h recalls were reported the day before the blood sample was drawn, and this could have affected the associations between diet and metabolomic variables. Moreover, our results need to be confirmed in other populations. At the same time, the analyses and predictions were conducted assessing total dietary fibre, but dietary fibre represents a heterogeneous group with large variations in physiochemical–physiological functions of importance to health.<sup>62,63</sup> In addition, although hs-CRP and increased blood pressure are well established cardiometabolic risk factors, our study design did not allow us to evaluate different rates of cardiovascular events for participants grouped according to clinical variables and metabolites. Likewise, the cumulative effects of dietary fibre intake on cardiometabolic risk factors cannot be assessed rigorously because 12 months is not a sufficient follow-up time, and previous dietary habits of the participants were unknown. More studies are needed to provide external validation of the proposed biomarkers and better characterize their differential production.

In conclusion, 2,6-DHBA and indolepropionic acid may represent a new set of biomarkers that reflect diet and host/gut microbiota interactions relevant to the cardiometabolic effects



of dietary fibre. Biomarkers reflecting the interactions between specific food components (dietary fibre) and host/gut microbiota may represent a novel measure to guide tailored diets for improved cardiometabolic health.

## Abbreviations

BMI	Body mass index
DCH-NG	Diet, Cancer and Health – Next Generations
FDR	False discovery rate
hsCRP	High sensitivity-C reactive protein
ICC	Intraclass-correlation coefficient
LDL-C	LDL cholesterol
MGM	Mixed Graphical Models
24-HDR	24 h dietary recall
2,6-DHBA	2,6-Dihydroxybenzoic acid
3,4-DHBA	3,4-Dihydroxybenzoic acid

## Author contributions

TM, AUC and CAL designed the research; RL, JH, and AT conceived and conceptualized the MAX-study; ARH, JH and AT collected the data of the DCH-NG cohort and the MAX-study; TM, AUC, RZR, ASP, AM, MC and CAL conducted the research; NET, RGD, TM, MMH and AUC conducted the metabolomic analysis; TM, AUC, RZR, ASP, AM, MCC and RZR performed statistical analysis; TM and AUC wrote the first draft of the manuscript; RZR, ARH, AT, GR, RL, JH and CAL provided critical revision. All authors read and approved the final manuscript.

## Data availability

Data may be available upon request to the Danish Cancer Society (contact: dchdata@cancer.dk).

## Conflicts of interest

There are no conflicts of interest to declare.

## Acknowledgements

The authors would like to thank all the participants of the study and the staff involved in the DCH-NG MAX-study. Tomás Meroño, Andrea Unión Caballero and Cristina Andrés-Lacueva thank MINECO PID2020-114921RB-C21, and CIBERFES (CB16/10/00269 from ISCIII). Cristina Andrés-Lacueva also thanks the award ICREA Academia 2018. Andrea Unión Caballero would like to thank the FPU 2020 contract financed by the Ministry of Universities (FPU20/05519). This work was accomplished as part of the DiGuMet project “Diet × gut microbiome-based metabolotypes to determine cardio-metabolic risk and tailor

intervention strategies for improved health” supported within the European Joint Programming Initiative “A Healthy Diet for a Healthy Life” (<https://www.healthydietforhealthylife.eu>), granted by Ministerio de Economía, Industria y Competitividad (MINECO) (Spain, PCIN-2017-076) and ERA-Net cofund on Intestinal Microbiomics (ERAHDHL INTIMICJPI HDHL). Project AC19/00111, funded by Instituto de Salud Carlos III and co-funded by European Regional Development Fund “A way to make Europe”. The work also received funding from CIBERFES, funded by the Instituto de Salud Carlos III and the Generalitat de Catalunya’s Agency AGAUR (2017SGR1546). The DCH-NG cohort was supported by the Danish Cancer Society, Knæk Cancer 2012 and Den A.P. Møllerske Støttefond (grant no. 10619). The establishment of the DCH-NG MAX study was partly funded by FORMAS (DNR 2016-00314). RZ-R was supported by the “Miguel Servet” program (CPII20/00009) from the Institute of Health Carlos III (co-funded by the European Social Fund (ESF) – ESF investing in your future). The contributions by RL was supported by Swedish Research Council grant no. 2019-01264 (Discovery and validation of novel biomarkers of gut microbiota, diet and their interactions associated with type 2 diabetes risk).

## References

- 1 J. M. Jones, CODEX-aligned dietary fiber definitions help to bridge the ‘fiber gap’, *Nutr. J.*, 2014, **13**(1), 34.
- 2 EFSA Panel on Dietetic Products, Nutrition, and Allergies (NDA), Scientific Opinion on Dietary Reference Values for carbohydrates and dietary fibre, *EFSA J.*, 2016, **8**(3).
- 3 A. M. Stephen, M. M. J. Champ, S. J. Cloran, M. Fleith, L. Van Lieshout, H. Mejbourn, *et al.*, Dietary fibre in Europe: current state of knowledge on definitions, sources, recommendations, intakes and relationships to health, *Nutr. Res. Rev.*, 2017, **30**(2), 149–190.
- 4 A. Reynolds, J. Mann, J. Cummings, N. Winter, E. Mete and L. Te Morenga, Carbohydrate quality and human health: a series of systematic reviews and meta-analyses, *Lancet*, 2019, **393**(10170), 434–445.
- 5 S. M. Murga-Garrido, Q. Hong, T. W. L. Cross, E. R. Hutchison, J. Han, S. P. Thomas, *et al.*, Gut microbiome variation modulates the effects of dietary fiber on host metabolism, *Microbiome*, 2021, **9**(1), 1–26.
- 6 P. Cronin, S. A. Joyce, P. W. O’toole and E. M. O’connor, Dietary Fibre Modulates the Gut Microbiota, *Nutrients*, 2021, **13**(5), 1655.
- 7 M. C. W. Myhrstad, H. Tunsjø, C. Charnock and V. H. Telle-Hansen, Dietary Fiber, Gut Microbiota, and Metabolic Regulation—Current Status in Human Randomized Trials, *Nutrients*, 2020, **12**(3), 859.
- 8 S. Shen, K. Prame Kumar, S. W. Wen, R. Shim, B. J. Wanrooy, D. Stanley, *et al.*, Deficiency of Dietary Fiber Modulates Gut Microbiota Composition, Neutrophil Recruitment and Worsens Experimental Colitis, *Front. Immunol.*, 2021, **12**, 193.



- 9 K. Makki, E. C. Deehan, J. Walter and F. Bäckhed, The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease, *Cell Host Microbe*, 2018, **23**(6), 705–715.
- 10 W. Ma, L. H. Nguyen, M. Song, D. D. Wang, E. A. Franzosa, Y. Cao, *et al.*, Dietary fiber intake, the gut microbiome, and chronic systemic inflammation in a cohort of adult men, *Genome Med.*, 2021, **13**(1), 1–13.
- 11 A. L. Barabási, G. Menichetti and J. Loscalzo, The unmapped chemical complexity of our diet, *Nat. Food*, 2019, **1**(1), 33–37.
- 12 E. M. Brouwer-Brolsma, L. Brennan, C. A. Drevon, H. Van Kranen, C. Manach, L. O. Dragsted, *et al.*, Combining traditional dietary assessment methods with novel metabolomics techniques: present efforts by the Food Biomarker Alliance, *Proc. Nutr. Soc.*, 2017, **76**(4), 619–627.
- 13 A. Scalbert, L. Brennan, C. Manach, C. Andres-Lacueva, L. O. Dragsted, J. Draper, *et al.*, The food metabolome: A window over dietary exposure, *Am. J. Clin. Nutr.*, 2014, **99**(6), 1286–1308.
- 14 C. Picó, F. Serra, A. M. Rodríguez, J. Keijer and A. Palou, Biomarkers of Nutrition and Health: New Tools for New Approaches, *Nutrients*, 2019, **11**(5), 1092.
- 15 A. Tebani and S. Bekri, Paving the way to precision nutrition through metabolomics, *Front. Nutr.*, 2019, **6**, 41.
- 16 A. B. Ross, A. Kamal-Eldin and P. Åman, Dietary alkylresorcinols: absorption, bioactivities, and possible use as biomarkers of whole-grain wheat- and rye-rich foods, *Nutr. Rev.*, 2004, **62**(3), 81–95.
- 17 R. Landberg, A. Kamal-Eldin, A. Andersson, B. Vessby and P. Åman, Alkylresorcinols as biomarkers of whole-grain wheat and rye intake: plasma concentration and intake estimated from dietary records, *Am. J. Clin. Nutr.*, 2008, **87**(4), 832–838.
- 18 A. M. Linko, K. S. Juntunen, H. M. Mykkänen and H. Adlercreutz, Whole-Grain Rye Bread Consumption by Women Correlates with Plasma Alkylresorcinols and Increases Their Concentration Compared with Low-Fiber Wheat Bread, *J. Nutr.*, 2005, **135**(3), 580–583.
- 19 I. Biskup, C. Kyrø, M. Marklund, A. Olsen, R. M. Van Dam, A. Tjønneland, *et al.*, Plasma alkylresorcinols, biomarkers of whole-grain wheat and rye intake, and risk of type 2 diabetes in Scandinavian men and women, *Am. J. Clin. Nutr.*, 2016, **104**(1), 88–96.
- 20 P. P. Söderholm, A. H. Koskela, J. E. Lundin, M. J. Tikkanen and H. C. Adlercreutz, Plasma pharmacokinetics of alkylresorcinol metabolites: new candidate biomarkers for whole-grain rye and wheat intake, *Am. J. Clin. Nutr.*, 2009, **90**(5), 1167–1171.
- 21 Y. Zhu, P. Wang, W. Sha and S. Sang, Urinary Biomarkers of Whole Grain Wheat Intake Identified by Non-targeted and Targeted Metabolomics Approaches, *Sci. Rep.*, 2016, **6**(1), 1–16. Available from: <https://www.nature.com/articles/srep36278>.
- 22 K. Hanhineva, P. Keski-Rahkonen, J. Lappi, K. Katina, J. Pekkinen, O. Savolainen, *et al.*, The postprandial plasma rye fingerprint includes benzoxazinoid-derived phenylacetamide sulfates, *J. Nutr.*, 2014, **144**(7), 1016–1022.
- 23 R. Landberg, K. Hanhineva, K. Tuohy, M. Garcia-Aloy, I. Biskup, R. Llorach, *et al.*, Biomarkers of cereal food intake, *Genes Nutr.*, 2019, **14**(1), 1–16.
- 24 P. Wang, J. Yang, A. Yerke and S. Sang, Avenacosides: Metabolism, and potential use as exposure biomarkers of oat intake, *Mol. Nutr. Food Res.*, 2017, **61**(7), 1700196, Available from: <https://pubmed.ncbi.nlm.nih.gov/28493602/>.
- 25 A. Johansson-Persson, T. Barri, M. Ulmius, G. Önning and L. O. Dragsted, LC-QTOF/MS metabolomic profiles in human plasma after a 5-week high dietary fiber intake, *Anal. Bioanal. Chem.*, 2013, **405**(14), 4799–4809.
- 26 K. E. N. Petersen, J. Halkjær, S. Loft, A. Tjønneland and A. Olsen, Cohort profile and representativeness of participants in the Diet, Cancer and Health—Next Generations cohort study, *Eur. J. Epidemiol.*, 2022, **37**(1), 117–127.
- 27 A. Tjønneland, A. Olsen, K. Boll, C. Stripp, J. Christensen, G. Engholm, *et al.*, Study design, exposure variables, and socioeconomic determinants of participation in Diet, Cancer and Health: A population-based prospective cohort study of 57,053 men and women in Denmark, *Scand. J. Public Health*, 2007, **35**(4), 432–441.
- 28 K. E. N. Petersen, J. Halkjær, S. Loft, A. Tjønneland and A. Olsen, Cohort profile and representativeness of participants in the Diet, Cancer and Health-Next Generations cohort study, *Eur. J. Epidemiol.*, 2022, **37**(1), 117–127.
- 29 P. A. Wark, L. J. Hardie, G. S. Frost, N. A. Alwan, M. Carter, P. Elliott, *et al.*, Validity of an online 24 h recall tool (myfood24) for dietary assessment in population studies: Comparison with biomarkers and standard interviews, *BMC Med.*, 2018, **16**(1), 1–14.
- 30 F. Lanuza, R. Zamora-Ros, A. L. Rostgaard-Hansen, A. Tjønneland, R. Landberg, J. Halkjær and C. Andres-Lacueva, Descriptive analysis of dietary (poly)phenol intake in the subcohort MAX from DCH-NG: “Diet, Cancer and Health—Next Generations cohort”, *Eur. J. Nutr.*, 2022, **1–14**, 337–350.
- 31 S. Guglielmetti, S. Bernardi, C. Del Bo’, A. Cherubini, M. Porrini, G. Gargari, *et al.*, Effect of a polyphenol-rich dietary pattern on intestinal permeability and gut and blood microbiomics in older subjects: Study protocol of the MaPLE randomised controlled trial, *BMC Geriatr.*, 2020, **20**(1), 1–10.
- 32 R. González-Domínguez, O. Jáuregui, P. Mena, K. Hanhineva, F. J. Tinahones, D. Angelino, *et al.*, Quantifying the human diet in the crosstalk between nutrition and health by multi-targeted metabolomics of food and microbiota-derived metabolites, *Int. J. Obes.*, 2020, **44**(12), 2372–2381.
- 33 P. Castellano-Escuder, R. Gonzalez-Domnguez, F. Carmona-Pontaque, C. Andrés-Lacueva and A. Sanchez-Pla, POMASHiny: A user-friendly web-based workflow for metabolomics and proteomics data analysis, *PLoS Comput. Biol.*, 2021, **17**(7), e1009148.
- 34 J. T. Leek, W. E. Johnson, H. S. Parker, A. E. Jaffe and J. D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments, *Bioinformatics*, 2012, **28**(6), 882–883.



- 35 B. T. West, K. B. Welch, A. T. Galecki and B. W. Gillespie, *Linear Mixed Models: A Practical Guide Using Statistical Software*, Linear Mixed Models, 2022.
- 36 M. Altenbuchinger, A. Weihs, J. Quackenbush, H. J. Grabe and H. U. Zacharias, Gaussian and Mixed Graphical Models as (multi-)omics data analysis tools, *Biochim. Biophys. Acta, Gene Regul. Mech.*, 2020, **1863**(6), 194418.
- 37 L. Shi, J. A. Westerhuis, J. Rosén, R. Landberg and C. Brunius, Variable selection and validation in multi-variate modelling, *Bioinformatics*, 2019, **35**(6), 972–980.
- 38 L. F. Călinoiu and D. C. Vodnar, Whole Grains and Phenolic Acids: A Review on Bioactivity, Functionality, Health Benefits and Bioavailability, *Nutrients*, 2018, **10**(11), 1615.
- 39 Y. Zhu, P. Wang, W. Sha and S. Sang, Urinary Biomarkers of Whole Grain Wheat Intake Identified by Non-targeted and Targeted Metabolomics Approaches, *Sci. Rep.*, 2016, **6**(1), 1–16.
- 40 O. Kärkkäinen, M. A. Lankinen, M. Vitale, J. Jokkala, J. Leppänen, V. Koistinen, *et al.*, Diets rich in whole grains increase betainized compounds associated with glucose metabolism, *Am. J. Clin. Nutr.*, 2018, **108**(5), 971–979.
- 41 M. Tuomainen, O. Kärkkäinen, J. Leppänen, S. Auriola, M. Lehtonen, M. J. Savolainen, *et al.*, Quantitative assessment of betainized compounds and associations with dietary and metabolic biomarkers in the randomized study of the healthy Nordic diet (SYSDIET), *Am. J. Clin. Nutr.*, 2019, **110**(5), 1108–1118.
- 42 D. R. Jacobs, M. A. Pereira, K. Stumpf, J. J. Pins and H. Adlercreutz, Whole grain food intake elevates serum enterolactone, *Br. J. Nutr.*, 2002, **88**(2), 111–116.
- 43 G. De Simone, C. Balducci, G. Forloni, R. Pastorelli and L. Brunelli, Hippuric acid: Could become a barometer for frailty and geriatric syndromes?, *Ageing Res. Rev.*, 2021, **72**, 101466.
- 44 M. Ulaszewska, M. Garcia-Aloy, N. Vázquez-Manjarrez, M. Soria-Florido, R. Llorach, F. Mattivi and C. Manach, Food intake biomarkers for berries and grapes, *Genes Nutr.*, 2020, **15**, 17.
- 45 S. S. Heinzmann, I. J. Brown, Q. Chan, M. Bictash, M. E. Dumas, S. Kochhar, *et al.*, Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption, *Am. J. Clin. Nutr.*, 2010, **92**(2), 436–443.
- 46 B. O. Keller, B. T. F. Wu, S. S. J. Li, V. Monga and S. M. Innis, Hypaphorine is present in human milk in association with consumption of legumes, *J. Agric. Food Chem.*, 2013, **61**(31), 7654–7660.
- 47 J. Chun, J. Lee, L. Ye, J. Exler and R. R. Eitenmiller, Tocopherol and tocotrienol contents of raw and processed fruits and vegetables in the United States diet, *J. Food Compos. Anal.*, 2006, **19**(2–3), 196–204.
- 48 M. Tuomainen, J. Lindström, M. Lehtonen, S. Auriola, J. Pihlajamäki, M. Peltonen, *et al.*, Associations of serum indolepropionic acid, a gut microbiota metabolite, with type 2 diabetes and low-grade inflammation in high-risk individuals, *Nutr. Diabetes*, 2018, **8**(1), 1–5.
- 49 C. Vilaplana-Pérez, D. Auñón, L. A. García-Flores and A. Gil-Izquierdo, Hydroxytyrosol and Potential Uses in Cardiovascular Diseases, Cancer, and AIDS, *Front. Nutr.*, 2014, **1**, 18.
- 50 I. A. T. Van De Luitgaarden, J. W. J. Beulens, I. C. Schrieks, L. M. Kienecker, D. J. Touw, A. J. Van Ballegooijen, *et al.*, Urinary Ethyl Glucuronide Can Be Used as a Biomarker of Habitual Alcohol Consumption in the General Population, *J. Nutr.*, 2019, **149**(12), 2199–2205.
- 51 T. K. Koo and M. Y. Li, A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research, *J. Chiropr. Med.*, 2016, **15**(2), 155.
- 52 J. Feng, Y. Qian, Z. Zhou, S. Ertmer, E. I. Vivas, F. Lan, *et al.*, Polysaccharide utilization loci in *Bacteroides* determine population fitness and community-level interactions, *Cell Host Microbe*, 2022, **30**(2), 200–215, DOI: [10.1016/j.chom.2021.12.006](https://doi.org/10.1016/j.chom.2021.12.006). Available from.
- 53 A. Reynolds, J. Mann, J. Cummings, N. Winter, E. Mete and L. Te Morenga, Carbohydrate quality and human health: a series of systematic reviews and meta-analyses, *Lancet*, 2019, **393**(10170), 434–445.
- 54 K. Makki, E. C. Deehan, J. Walter and F. Bäckhed, The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease, *Cell Host Microbe*, 2018, **23**(6), 705–715.
- 55 D. Vojinovic, D. Radjabzadeh, A. Kurilshikov, N. Amin, C. Wijmenga, L. Franke, *et al.*, Relationship between gut microbiota and circulating metabolites in population-based cohorts, *Nat. Commun.*, 2019, **10**(1), 1–7, DOI: [10.1038/s41467-019-13721-1](https://doi.org/10.1038/s41467-019-13721-1). Available from.
- 56 L. Zhao, Q. Zhang, W. Ma, F. Tian, H. Shen and M. Zhou, A combination of quercetin and resveratrol reduces obesity in high-fat diet-fed rats by modulation of gut microbiota, *Food Funct.*, 2017, **8**(12), 4644–4656.
- 57 L. Zhang, Y. Ouyang, H. Li, L. Shen, Y. Ni, Q. Fang, *et al.*, Metabolic phenotypes and the gut microbiota in response to dietary resistant starch type 2 in normal-weight subjects: a randomized crossover trial, *Sci. Rep.*, 2019, **9**(1), 1–11.
- 58 A. Latorre-Pérez, M. Hernández, J. R. Iglesias, J. Morán, J. Pascual, M. Porcar, *et al.*, The Spanish gut microbiome reveals links between microorganisms and Mediterranean diet, *Sci. Rep.*, 2021, **11**(1), 1–12, DOI: [10.1038/s41598-021-01002-1](https://doi.org/10.1038/s41598-021-01002-1). Available from.
- 59 M. Knez, M. Ranic, J. C. R. Stangoulis and M. Glibetic, 2.20 - The Mineral Intake and Microbiota, *Compr. Gut Microbiota*, 2022, **2**, 230–242.
- 60 L. Li, S. Yan, S. Liu, P. Wang, W. Li, Y. Yi, *et al.*, In-depth insight into correlations between gut microbiota and dietary fiber elucidates a dietary causal relationship with host health, *Food Res. Int.*, 2023, **172**, 113133, DOI: [10.1016/j.foodres.2023.113133](https://doi.org/10.1016/j.foodres.2023.113133). Available from.
- 61 H. C. Wastyk, G. K. Fragiadakis, D. Perelman, D. Dahan, B. D. Merrill, F. B. Yu, *et al.*, Gut-microbiota-targeted diets modulate human immune status, *Cell*, 2021, **184**(16), 4137–4153. Available from: <https://pubmed.ncbi.nlm.nih.gov/34256014/>.



- 62 E. C. Deehan, C. Yang, M. E. Perez-Muñoz, N. K. Nguyen, C. C. Cheng, L. Triador, *et al.*, Precision Microbiome Modulation with Discrete Dietary Fiber Structures Directs Short-Chain Fatty Acid Production, *Cell Host Microbe*, 2020, **27**(3), 389–404.
- 63 N. T. Baxter, A. W. Schmidt, A. Venkataraman, K. S. Kim, C. Waldron and T. M. Schmidt, Dynamics of Human Gut Microbiota and Short-Chain Fatty Acids in Response to Dietary Interventions with Three Fermentable Fibers, *mBio*, 2019, **10**(1), e02566.

