Metabolomics biomarkers for diet and adiposity

Doctoral thesis

to obtain a doctorate (PhD)

from the Faculty of Medicine

of the University of Bonn

Samuel Mutisya Muli

from Kitui, Kenya

2025

Written with authorization of

the Faculty of Medicine of the University of Bonn

First reviewer: Prof. Dr. Ute Nöthlings

Second reviewer: Prof. Dr. Matthias Schmid

Day of oral examination: January 9, 2025

From the Institute of Nutrition and Food Sciences

In cherished memory of my father

Table of contents

List of Abbreviations

AS	Added sugars
BMI	Body mass index
DONALD	Dortmund Nutritional and Anthropometric Longitudinal Designed
FFQ	Food frequency questionnaire
LASSO	Least absolute shrinkage and selection operator
LNCSB	Low- and no-calorie sweetened beverages
PCA	Principal component analysis
SBs	Sweetened beverages
SSB	Sugar-sweetened beverages
UPF	Ultra processed foods
WC	Waist circumference
WHO	World Health Organization
24h-DRs	24h-dietary recalls
%BF	body fat percentage

1 Abstract

Diet is an important determinant of health and well-being. Epidemiologic studies link higher habitual intake of sweetened beverages (SBs), sweet and fatty snacks, and the broad range of ultra-processed foods (UPF) with an increased risk of obesity. Poor nutritional profiles, higher caloric intake, and energy imbalance are some of the proposed mechanisms, but other biological pathways underlying diet-related weight gain and regulation are not fully defined. Human metabolome provides a rich resource for understanding metabolic alterations associated with diet. Here, we conduct a literature review on biomarkers of SBs (study 1); investigate the metabolomic signatures of SBs and added sugar intake in children, adolescents, and young adults and their association with adiposity measures (study 2); investigate the reproducibility of urine biomarkers of sweet and fatty snacks across two independent cohorts (study 3); and investigate the metabolomic profiles of UPF intake in adolescents and young adults and their association with adiposity (study 4).

In study 1, we conducted a systematic review of the literature on biomarkers of SBs and their levels of validity. In study 2, we used 3 data sets across 3 age groups: children (3.0–10.3 y), adolescents (14.9–18.4 y), and young adults (18.0–21.9 y), from the DONALD cohort study. In study 3, we included the previously defined sample of children and children from an external cohort, the IDEFICS/I.Family cohort. In study 4, we included the adolescent and young adult analytic samples defined in study 2. We used untargeted metabolomics in urine and plasma across all studies and additionally conducted lipidomics on plasma. We applied multiple machine learning methods because of the high-dimensional data: the random forest, partial least squares, and LASSO for joint metabolite selection (study 2 and 3); particle swarm optimization and extreme gradient boosting for investigating metabolite data missing mechanisms (study 4); and robust sparse PCA for deriving metabolite patterns (study 4). We used linear and mixed effects for covariate adjustments (study 2-4).

We identified metabolomic signatures of SBs, added sugar, sweet and fatty snacks, and UPF intake in young individuals. Some of these metabolomic changes were related to adiposity measures and may be important research targets for better understanding of the mechanisms through which these foods contribute to weight gain and adiposity.

2 Introduction and aims with references

2.1 Introduction

Diet is a well-established modifiable lifestyle factor for health and well-being. Healthy dietary habits from childhood can improve long-term health outcomes by reducing the risk of diet-related diseases (Neuhouser et al., 2023; Lane et al., 2024). A major concern regarding the modern dietary patterns is the regular and often excessive consumption of sweetened beverages (SBs), which broadly include sugar-sweetened beverages (SSB) and low- and no-calorie sweetened beverages (LNCSB). SSBs are a major public health nutrition issue because of the added sugars (AS), which contribute to excessive caloric intake and overall poor diet (Marriott et al. 2019; Endy et al. 2024). Individuals with higher intake of SSBs also tend to consume other foods of poorer quality compared with non-SSB consumers (Doherty, Lacko, & Popkin, 2021), suggesting the need for a broader focus on diet quality beyond SSBs. This evidence is consistent. For example, dietary patterns across the primary-secondary school transition showed that SBs are habitually consumed alongside other discretionary and energy-dense foods e.g., cakes, pastries, crisps, and savory snacks (Peral-Suárez et al., 2024).

These foods are main subgroups of ultra-processed foods (UPF) as defined by the NOVA food classification system, which broadly includes soft drinks, flavored yoghurts, sweet and savory snacks, cakes and confectionery, chocolates, pizza, chips, mass-produced breads and buns, breakfast cereals, sausages, processed and reconstituted meats, margarine and spreads, pasta, noodles and a wide range of ready-to-eat or-heat convenient food products, primarily or entirely derived from food ingredients and additives of exclusive industrial use (Monteiro et al., 2019). Some of the factors that contribute to their widespread popularity and consumption include convenience, accessibility, affordability, and a long shelf-life (Monteiro et al., 2019). There is widespread debate on UPFs, unsurprisingly on the degree of processing in relation to healthy food: “Even we are confused,” admitted an interdisciplinary consortium drawn from nutrition, food technology, industry, policy, and civil society (Sadler et al., 2022).

One point is evident, though. UPFs as currently defined are the fastest growing segment of the global and largely industrialized food systems (Northcott et al., 2023); a growth consistent with an incremental shift in dietary patterns from traditional to more Western-

style diets (Beslay et al., 2020). In line, the global prevalence of obesity has also been on a consistent trajectory with this worldwide shift toward more processed diets (Monteiro et al. 2019; Baker et al. 2020). For example, in 2022, the World Health Organization (WHO) estimated that over 390 million children and adolescents (aged 5–19) were overweight, with 160 million classified as obese (WHO, 2024). While the role of dietary guidelines in reducing the global burden of obesity is certainly clear, the specific foods and mechanisms that more significantly explain weight gain and adiposity beyond the traditional energy balance models are widely debated (Hall et al., 2022). This is compounded by another long-standing problem: challenges in dietary assessment.

So far, evidence from nutritional epidemiologic studies is based on self-reported intakes using food frequency questionnaires (FFQ), 24h-dietary recalls (24h-DRs), and food records, which are prone to random and systematic errors. This constitutes a significant challenge that, an increasing number of studies are dedicated to measurement bias and correction of errors in self-reported dietary intakes (Keogh et al., 2020; Hu et al., 2022). Foods traditionally believed to contribute more to obesity risk such as SBs are also likely to be underreported because of social desirability bias (Knox et al., 2020). Indeed, an earlier study on the tendency to report and the portion size if reported, showed that low energy reporters were more likely to report smaller sizes across several food groups, including soft drinks (Krebs-Smith et al., 2000). It is, therefore, difficult to differentiate true null associations between diet and health status from those due to misreporting and measurement errors. For example, there are inconsistent findings on the associations between SBs and UPF and health outcomes (McGlynn et al., 2022; Steffen et al., 2023; Kermani-Alghoraishi et al., 2024).

Beyond the diet-disease associations, assessment of diet using two different instruments on the same individual also shows inconsistencies in self-rated diet quality, mainly due to subjective interpretation of dietary questions, framing effects of these questions, and self-awareness of nutritional status (Nouve, Zhao, & Zheng, 2024). To address some of these challenges, some studies combine data from FFQs and 24-h DRs with unbiased biomarkers, such as those for nutrients, using regression calibration approaches (Looman et al., 2019), which has been shown to provide more accurate assessments of the associations of habitual intakes with health outcomes (Looman et al., 2019; Huang et al.,

2022). This approach is promising for improving dietary assessment, but few validated biomarkers of intake exist.

Metabolomics techniques for profiling small molecules in biosamples have emerged as promising tools in dietary biomarker research. The human metabolome comprises the endogenous metabolome, consisting of metabolites from own cells; the food metabolome, which includes nutrients and end products of biotransformation of food substrates; and xenobiotics from drugs and chemicals, environmental and other exposures (Scalbert et al., 2014). There is a rapidly growing interest in metabolomics given that metabolites represent the most proximal layer to phenotypes; hence, they reflect phenotypic variations due to a disease state, diet, or environmental exposures (Scalbert et al., 2014; Auwerx et al., 2023).

A recent large cohort study identified serum and 24-h urine metabolites of red meat and the combination of red and processed meat, and applied calibration approaches (like the aforementioned approaches for single nutrients) to correct systematic measurement errors in their habitual intakes. This achieved a better estimation of the association between these intakes and cardiovascular disease, cancer, and diabetes incidence (Zheng et al., 2022). Urinary proline betaine has also been used to correct self-reported usual intakes of citrus fruits (Hu et al., 2024). Regarding SBs, a study comparing self-reports (FFQ and 24h-DRs) with urinary biomarkers of LNCSB intake demonstrated that these dietary instruments underestimated overall dietary exposures to low calorie sweeteners (Buso et al., 2024). These studies demonstrate significant progress in the application of metabolomics to improve nutrition research.

A particular strength of metabolomics is that, apart from identifying biomarkers of dietary intake, the molecular fingerprints of intake also provide biological insights into the mechanisms through which these dietary exposures influence metabolism and health. For example, phosphatidylcholine and lysophospholipid have been proposed as modulating pathways linking SSB intake and obesity risk (Zhou et al., 2020) and lipids and glutamate metabolism as some of the biological links between SSB and dairy intakes and obesity and type 2 diabetes (Parnell et al., 2021).

We propose to use untargeted metabolomics approaches to investigate the biological mechanisms through which SBs and AS, sweet and fatty snacks, and UPF influence metabolism and their association with adiposity. This is particularly important given the overwhelming epidemiologic evidence linking habitual intake of UPF or its subgroups with a higher risk for obesity (Beslay et al., 2020; Endy et al., 2024; Lane et al., 2024). To achieve this, we use data from the DONALD cohort study, which uses 3-d weighed dietary records and provides rich data on intake at the food item level rather than food group level, characteristic of FFQs. This granular, food item level approach provides a better separation of SBs into SSB and LNCSB, a better estimation of AS from all food items consumed, and a more precise categorization of the UPF. The regular urine, diet, and adiposity assessments further enable this study to investigate metabolomic changes associated with both short-term and long-term intake of these foods, and their relationship with adiposity phenotypes.

2.2 Aims

The overarching objective was to gain a better understanding of urine- and plasma-based metabolomic changes of short-term and long-term intakes of SBs, sweet and fatty snacks, and UPF; and to investigate whether these metabolic alterations are related to adiposity. Therefore, we defined the following specific objectives:

1. To perform a systematic review of metabolomic biomarkers of SSB and LNCSB intake (see 3.1, publication 1).
2. To identify the plasma and urine metabolomic signatures of SBs and AS intake across various age groups various age groups (children, adolescents, and young adults) and to evaluate their relationship with BMI, %BF, and WC (see 3.2, publication 2).
3. To identify and replicate urine metabolites associated with short-term and habitual intake of sweet and fatty snacks in European children and adolescents (see 3.3, publication 3).
4. To investigate the associations of UPF intake with urine and plasma metabolomic profiles in adolescents and young adults (see 3.4, publication 4), and their association with BMI, %BF, and WC (see 3.4, extended analyses on publication 4).

2.3 References

Auwerx C, Sadler M, Woh T, Reymond A, Kutalik Z, Porcu E. Exploiting the mediating role of the metabolome to unravel transcript-to-phenotype associations. Elife. 2023; 12:e81097

Baker P, Machado P, Santos T, Sievert K, Backholer K, Hadjikakou M, et al. Ultra-processed foods and the nutrition transition: Global, regional and national trends, food systems transformations and political economy drivers. Obes Rev 2020; 21: e13126

Beslay M, Srour B, Méjean C, Allès B, Fiolet T, Debras C, et al. Ultra-processed food intake in association with BMI change and risk of overweight and obesity: A prospective analysis of the French NutriNet-Santé cohort. PLoS Med 2020; 17: e1003256

Buso M, Boshuizen H, Naomi N, Maho W, Diepeveen-de Bruin M, Balvers M, et al. Relative validity of habitual sugar and low/no-calorie sweetener consumption assessed by food frequency questionnaire, multiple 24-h dietary recalls and urinary biomarkers: an observational study within the SWEET project. Am J Clin Nutr 2024; 119:546-559

Doherty A, Lacko A, Popkin B. Sugar-sweetened beverage (SSB) consumption is associated with lower quality of the non-SSB diet in US adolescents and young adults. Am J Clin Nutr 2021; 113: 657-664

Endy EJ, Yi S, Steffen BT, Shikany JM, Jacobs DR, Goins RK, et al. Added sugar intake is associated with weight gain and risk of developing obesity over 30 years: The CARDIA study. Nutr Metab Cardiovasc Dis 2024; 34: 466-474.

Hall K, Farooqi I, Friedman J, Klein S, Loos R, Mangelsdorf D, et al. The energy balance model of obesity: beyond calories in, calories out. Am J Clin Nutr 2022; 115:1243-1254.

Hu Y, Wang M, Willett WC, Stampfer M, Liang L, Hu FB, et al. Calibration of citrus intake assessed by food frequency questionnaires using urinary proline betaine in an observational study setting. Am J Clin Nutr 2024; 120: 178-186

Hu Y, Tang D, Yang F, Dai S, Xiao X, Zhao X. The impacts of measurement errors on a dietary pattern analyses: a simulation study based on dietary data from the China Multi-Ethnic Cohort (CMEC) study. Am J Clin Nutr 2022; 116: 523-530

Huang Y, Zheng C, Tinker LF, Neuhouser ML, Prentice RL. Biomarker-based methods and study designs to calibrate dietary intake for assessing diet–disease associations. J Nutr 2022; 152: 899-906

Keogh RH, Shaw PA, Gustafson P, Carroll RJ, Deffner V, Dodd KW, et al. STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: Part 1—Basic theory and simple methods of adjustment. Stat Med 2020; 39: 2197-2231

Kermani-Alghoraishi M, Behrouzi A, Hassannejad R, Sarrafzadegan N, Nouri F, Boshatam M, et al. Ultra-processed food consumption and cardiovascular events rate: An analysis from Isfahan Cohort Study (ICS). Nutr Metab Cardiovasc Dis 2024; 34: 1438-1447

Knox MA, Oddo VM, Walkinshaw LP, Jones-Smith J. Is the public sweet on sugary beverages? Social desirability bias and sweetened beverage taxes. Econ Hum Biol 2020; 38: 100886

Krebs-Smith SM, Graubard BI, Kahle LL, Subar AF, Cleveland LE, Ballard-Barbash R. Low energy reporters vs others: a comparison of reported food intakes. Eur J Clin Nutr 2000; 54: 281-287

Lane MM, Gamage E, Du S, Ashtree DN, McGuinness AJ, Gauci S, et al. Ultra-processed food exposure and adverse health outcomes: umbrella review of epidemiological meta-analyses. BMJ 2024; 384: e077310

Looman M, Boshuizen HC, Feskens EJ, Geelen A. Using enhanced regression calibration to combine dietary intake estimates from 24 h recall and FFQ reduces bias in diet–disease associations. Public Health Nutr 2019; 22: 2738-2746

Marriott BP, Hunt KJ, Malek AM, Newman JC. Trends in intake of energy and total sugar from sugar-sweetened beverages in the United States among children and adults, NHANES 2003–2016. Nutrients 2019; 11: 2004

McGlynn ND, Khan TA, Wang L, Zhang R, Chiavaroli L, Au-Yeung F, et al. Association of low- and no-calorie sweetened beverages as a replacement for sugar-sweetened

beverages with body weight and cardiometabolic risk: a systematic review and meta-analysis. JAMA Netw Open 2022; 5: e222092-e92

Monteiro CA, Cannon G, Levy RB, Moubarac JC, Louzada M, Rauber F, et al. Ultra-processed foods: what they are and how to identify them. Public Health Nutr 2019; 22:936-941

Neuhouser ML, Prentice RL, Tinker LF, Lampe JW. Enhancing capacity for food and nutrient intake assessment in population sciences research. Annu Rev Public Health 2023; 44: 37-54

Northcott T, Lawrence M, Parker C, Baker P. Ecological regulation for healthy and sustainable food systems: responding to the global rise of ultra-processed foods. Agric Human Values 2023; 40: 1333-1358

Nouve Y, Zhao S, Zheng Y. Decoding the misperception: Exploring measurement error in self-rated assessments of diet quality. Food Qual Prefer 2024; 120:105234

Parnell LD, Noel SE, Bhupathiraju SN, Smith CE, Haslam DE, Zhang X, et al. Metabolite patterns link diet, obesity, and type 2 diabetes in a Hispanic population. Metabolomics. 2021; 17: 88

Peral-Suárez Á, Haycraft E, Blyth F, Holley CE, Pearson N. Dietary habits across the primary-secondary school transition: A systematic review. Appetite. 2024; 201:107612

Sadler C, Grassby T, Hart K, Raats M, Sokolovic M, Timotijevic L. “Even we are confused”: A thematic analysis of professionals’ perceptions of processed foods and challenges forcommunication. Front Nutr 2022; 9: 826162

Scalbert A, Brennan L, Manach C, Andres-Lacueva C, Dragsted LO, Draper J, et al. The food metabolome: a window over dietary exposure. Am J Clin Nutr 2014; 99:1286-1308

Steffen BT, Jacobs DR, Yi SY, Lees SJ, Shikany JM, Terry JG, et al. Long-term aspartame and saccharin intakes are related to greater volumes of visceral, intermuscular, and subcutaneous adipose tissue: the CARDIA study. Int J Obes 2023; 47: 939-947

WHO, 2024. Obesity and overweight. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight (date accessed: 15.09.2024)

Zheng C, Pettinger M, Gowda GAN, Lampe JW, Raftery D, Tinker LF, et al. Biomarker-calibrated red and combined red and processed meat intakes with chronic disease risk in a cohort of postmenopausal women. J Nutr 2022; 152: 1711-1720

Zhou B, Ichikawa R, Parnell LD, Noel SE, Zhang X, Bhupathiraju SN, et al. Metabolomic links between sugar-Sweetened beverage intake and obesity. J Obes 2020; 2020:7154738

3 Publications

3.1 Publication 1

A systematic review of metabolomic biomarkers for the intake of sugar-sweetened and low-calorie sweetened beverages. https://doi.org/10.3390/metabo11080546

metabolites

MDPI

Review

A Systematic Review of Metabolomic Biomarkers for the Intake of Sugar-Sweetened and Low-Calorie Sweetened Beverages

Samuel Muli [1,*], Jantje Goerdten [2], Kolade Oluwagbemigun [1], Anna Floegel [2], Matthias Schmid [3] and Ute Nöthlings [1]

[1] Nutritional Epidemiology, Department of Nutrition and Food Sciences, University of Bonn, Friedrich-Hirzebruch-Allee 7, 53115 Bonn, Germany; koluwagb@uni-bonn.de (K.O.); noethlings@uni-bonn.de (U.N.)
[2] Department of Epidemiological Methods and Etiological Research, Leibniz Institute for Prevention Research and Epidemiology—BIPS, Achterstr. 30, 28359 Bremen, Germany; goerdten@leibniz-bips.de (J.G.); floegel@leibniz-bips.de (A.F.)
[3] Institute for Medical Biometry, Informatics and Epidemiology (IMBIE), University Hospital Bonn, Venusberg Campus 1, 53127 Bonn, Germany; matthias.schmid@imbie.uni-bonn.de
* Correspondence: smuli@uni-bonn.de; Tel.: +49-(0)-228-73-3765

Abstract: Intake of added sugars (AS) is challenging to assess compared with total dietary sugar because of the lack of reliable assessment methods. The reliance on self-reported dietary data in observational studies is often cited as biased, with evidence of AS intake in relation to health outcomes rated as low to moderate quality. Sugar-sweetened beverages (SSBs) are a major source of AS. A regular and high intake of SSBs is associated with an overall poor diet, weight gain, and cardiometabolic risks. An elevated intake of low-calorie sweetened beverages (LCSBs), often regarded as healthier alternatives to SSBs, is also increasingly associated with increased risk for metabolic dysfunction. In this review, we systematically collate evidence and provide perspectives on the use of metabolomics for the discovery of candidate biomarkers associated with the intake of SSBs and LCSBs. We searched the Medline, Embase, Scopus, and Web of Science databases until the end of December 2020. Seventeen articles fulfilled our inclusion criteria. We evaluated specificity and validity of the identified biomarkers following Guidelines for Biomarker of Food Intake Reviews (BFIRev). We report that the ^{13}C:^{12}C carbon isotope ratio ($\delta^{13}C$), particularly, the $\delta^{13}C$ of alanine is the most robust, sensitive, and specific biomarker of SSBs intake. Acesulfame-K, saccharin, sucralose, cyclamate, and steviol glucuronide showed moderate validity for predicting the short-term intake of LCSBs. More evidence is required to evaluate the validity of other panels of metabolites associated with the intake of SSBs.

Keywords: sugar-sweetened beverages; low-calorie sweetened beverages; metabolomics; biomarkers

Citation: Muli, S.; Goerdten, J.; Oluwagbemigun, K.; Floegel, A.; Schmid, M.; Nöthlings, U. A Systematic Review of Metabolomic Biomarkers for the Intake of Sugar-Sweetened and Low-Calorie Sweetened Beverages. *Metabolites* **2021**, *11*, 546. https://doi.org/10.3390/metabo11080546

Academic Editor: Amedeo Lonardo

Received: 6 July 2021
Accepted: 14 August 2021
Published: 19 August 2021

1. Introduction

Added sugar (AS) refers to sugars, syrups, or caloric sweeteners added to foods during preparation, processing in the industry, or by consumers at the table [1]. A high intake of AS is a public health concern, because of its associated health risks. The World Health Organization (WHO) recommends less than 10% of the total daily energy intake from free sugars, which includes AS and sugars naturally present in honey, syrups, fruit concentrates, and juices [2]. In reference to a total energy intake of 2000 kcal per day, the WHO recommendation corresponds to 50 g of free sugars [3]. Despite the imprecise definition of AS and free sugars in epidemiologic studies, there is consensus that consumption significantly exceeds WHO recommendations. In a German cohort study, the median intake of AS ranged between 11.6% and 13.3% and free sugars between 15.2% to 17.5% in children and adolescents aged 3 to 18 years [4]. In the USA, a national survey reported a mean adjusted estimate of AS intake in children aged 2–18 years as 14% of their daily energy intake [5].

A major source of dietary AS is sugar-sweetened beverages (SSBs). A regular and high intake of SSBs is associated with overall poor diet quality [6], weight gain and progression of obesity [7], increased risk for diabetes [8], cardiovascular diseases [9–11], and a low-grade inflammatory state [12,13]. Low-calorie sweetened beverages (LCSBs), which contain non-nutritive sweeteners, are commonly marketed as healthier alternatives to SSBs [14]. However, emerging evidence from observational studies suggests their inverse association with cardiometabolic health [15], including risk for ischemic stroke and all cause dementia [16], impaired insulin sensitivity in healthy individuals [17], and increased risk for cardiovascular diseases [18]. A causal link between SSBs/LCSBs intake and negative health effects is yet to be established.

As all consumed foods like SSBs/LCSBs are metabolized, their metabolites could be a window to their intake and may also improve our understanding of the causal link with the aforementioned health conditions. This is especially important because varying opinions persist about evidence from self-reported dietary assessment tools, such as dietary food records, 24-h dietary recalls, and food frequency questionnaires (FFQs). These instruments are often cited as having inherent biases of recall and misreporting, which may lead to incorrect estimations of the associations between intake and health outcomes [19]. Evidence from studies suggest that foods considered socially undesirable, such as those high in AS like SSBs, are mostly underreported [20,21]. In part, such challenges have continued to fuel the debate on the validity of the associations between the intake of SSBs/LCSBs and health risks. This potentially undermines public health messages that urge the public to reduce the intake of AS and SSBs. Objective biomarkers for the dietary intake of SSBs/LCSBs could aid in overcoming this longstanding challenge by complementing the existing dietary instruments to strengthen the evidence on connection between intake and health status [22].

The discovery and validation of biomarkers of SSB and LCSB intake remains a high-priority research area, with rapidly growing evidence of dietary signatures in blood, urine, fingernails, hair, and other human tissues [23]. Some biomarkers have been proposed through targeted metabolomics methods, and have been validated in small, controlled feeding studies. However, untargeted methods of biosamples following exposure to SSBs and LCSBs have also produced panels of novel metabolites that need further investigation and validation. Therefore, this study collates the latest evidence from studies applying metabolomics methods for the discovery of candidate biomarkers associated with the intake of SSBs and LCSBs.

2. Materials and Methods

2.1. Literature Search

To identify the biomarkers of food intake (BFI) for SSBs and LCSBs, an extensive literature search was conducted following the Guidelines for Biomarker of Food Intake Reviews (BFIRev) [24] and the PRISMA statement for systematic reviews [25], whenever meaningful. We registered the review project with the OSF Registries (DOI: 10.17605/OSF.IO/97VFY, https://osf.io/2pvr3/, accessed on 19 August 2021). We comprehensively searched four electronic databases, Medline, Embase (in OVID SP), Scopus, and Web of Science, using the following search terms, adapted appropriately to each database: (sugar*sweet*beverage* OR SSB* OR beverage* OR added sugar* OR caloric*sweet* OR soda* OR diet*beverage* OR soft drink* OR low*calorie*sweet*beverage* OR LCSB* OR artificial*sweet*beverage* OR ASB* OR fruit flavored drink* OR carbonated drink* OR juice*) AND (biomarker* OR marker* OR metabolite* OR metabolom* OR biomonitor* OR biosignature* OR bioavailability) AND (intake OR diet OR dietary pattern* OR dietary habit* OR eating pattern* OR food* OR meal* OR nutrition*assessment OR nutrition* survey*) AND (plasma OR urin* OR serum OR blood OR hair). The search was limited to papers published on human studies and in English, from inception dates until the end of December 2020. Studies on animal models were excluded. We used EndNote (version X9) and Rayyan QCRI programs for reference management and abstract screening, respectively. Two independent reviewers

(S.M. and J.G.) screened all the titles and abstracts, and conducted full text reading and subsequent data extraction.

2.2. Evaluation of Specificity of Identified Biomarkers

We performed an extensive literature search, as recommended in the BFIRev guidelines [24], to evaluate the specificity of the identified candidate biomarkers. In a second search step, we evaluated the specificity of the candidate biomarkers in the Human Metabolome Database (HMDB), the Food Database (FooDB), and the Phenol-Explorer. If the reported compound was identified as a biomarker for non-SSBs and non-LCSBs food items, such a compound would be removed from further selection. Next, to confirm if the identified potential biomarkers were detected in other foods, further literature search was performed: ("name and synonyms of biomarker candidate") AND (biomarker* OR marker* OR metabolite* OR metabolom* OR biomonitor* OR biosignature*), which was executed on the Google Scholar search engine. Compounds present in other foods were determined as lacking specificity for SSBs or LCSBs.

2.3. Evaluating of Validity of Biomarkers

We adopted the framework proposed by Dragsted et al. [26] to assess the validity of the identified biomarkers of SSBs and LCSBs. This framework provides eight groups of validity criteria for assessing the validation and application of BFIs, namely, plausibility, dose−response relationship, time−response (single-meal time response and repeated intakes), robustness, reliability, stability, analytical performance, and reproducibility. In total, the validity of the candidate biomarkers was assessed by answering nine questions, with either a yes, no, or uncertain/unknown. Selected biomarkers were then graded, with the scores reflecting the current validity rating of the biomarker as informed by available evidence.

2.4. Evaluating Quality of Evidence

Because of the lack of standard validated tools for evaluating the quality of evidence of the metabolomics studies, we applied two assessment tools to assess the risk of bias and biomarker measurement characteristics. For quality assessment of the evidence (i.e., risk of bias and study quality), we applied the NutriGrade scoring system, which uses the Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) approach [27]. To evaluate the quality of the biomarker measurement, we applied the Biomarker-based Cross-sectional studies (BIOCROSS) evaluation tool, which is especially adapted for biomarker measurement, representing biosample and assay methods, laboratory measurement, and biomarker data models [28].

3. Results

The systematic literature search strategy yielded 1130 non-duplicated records from the four electronic databases. After abstract reading, full text reading for eligibility assessment, and secondary search, 17 studies were included [29–45], as shown in Figure 1.

Table 1 summarizes the study characteristics and candidate biomarkers identified. There were eight cross-sectional studies [29,30,33,34,36,38,39,44,45] and eight controlled intervention studies [31,32,35,37,40–43]. For comparison with dietary intake, dietary assessment tools such as 24-h recall, 4-day and 7-day food records, and FFQS were used. Four studies were conducted in children and adolescents [35,38,39,44], two in postmenopausal women [41,42], one in predominantly obese population [31], and one study in an inpatient hospital setting [40]. Two studies used an untargeted metabolomics approach [33,39], while the rest used a targeted metabolomics approach.

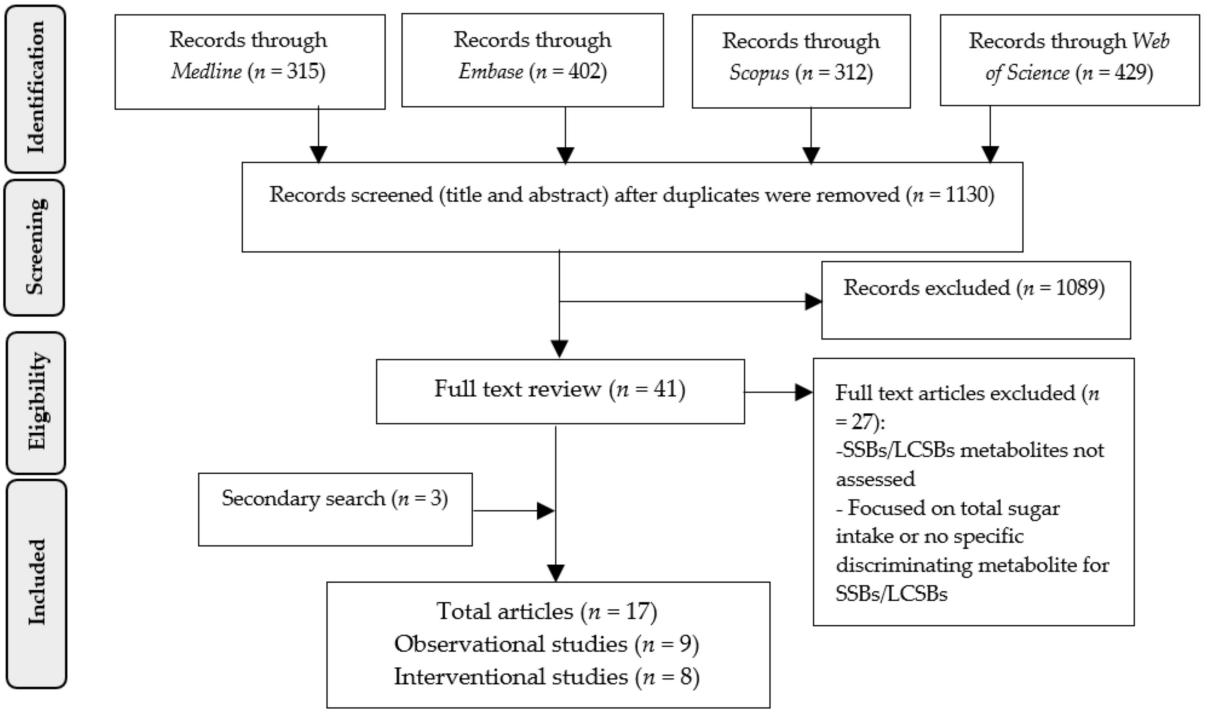

Figure 1. PRISMA flowchart on the screening process and selection of papers reporting biomarkers of SSBs and LCSBs as of December 2020.

Table 1. Studies on the association between consumption and potential candidate biomarkers for sugar- and low-calorie sweetened beverages.

Study, Country, [Reference]	Number of Participants	Age Range (Years)	Dietary Assessment Method	Sample Type	Chemical Analytic Method	Analytic Approach	Candidate Biomarker of Food Intake/Metabolite
Davy et al. 2017, USA [31]	301	≥18	24-h recall (×3)	Fasting fingerstick blood	NA-SIMS	Targeted	$\delta^{13}C$
Choy et al. 2013, USA [29]	68	14–79	24-h recall (×4)	Red blood cells, hair	GC-IRMS	Targeted	$\delta^{13}C$–alanine
Davy et al. 2011, USA [30]	60	≥21	4-d DR	fingerstick blood	NA-SIMS	Targeted	$\delta^{13}C$
Fakhouri et al. 2014, USA [32]	144	25–79	24-h recall (×2)	Serum, after 8-h fast	IRMS	Targeted	$\delta^{13}C$
Hedrick et al. 2016, USA [34]	216	≥18	24-h recall (×3)	Fasting fingerstick blood	IRMS	Targeted	$\delta^{13}C$
Nash et al. 2014, USA [45]	68	14–79	24-h recall (×4)	Red blood cells, plasma, hair	IRMS	Targeted	$\delta^{13}C$
Votruba et al. 2019, USA [40]	32	46.2 (10.5) [a]	7-d DR	Plasma, hair, Red blood cells	IRMS	Targeted	$\delta^{13}C$
Liu et al. 2018, USA [35]	33	12–18	24-h recall (×8)	Fasting fingerstick blood	NA-SIMS	Targeted	$\delta^{13}C$
Yun et al. 2018, USA [41] **	153	75 (4) [a]	4-d DR	Serum	IRMS	Targeted	$\delta^{13}C$
Yun et al. 2020, USA [42]	145	75 (73, 78) [b]	4-d DR	Serum AAs	GC-IRMS	Targeted	$\delta^{13}C$–alanine
MacDougall et al. 2018, USA [38]	126	6–11	24-h recall (×4)	Fingerstick blood	IRMS	Targeted	$\delta^{13}C$
Valenzuela et al. 2018, USA [44]	212	9–16	FFQ	Hair, Breath	GC-IRMS	Targeted	$\delta^{13}C$
Gibbons et al. 2015, Ireland [33]	565	≥18	4-d DR	Urine	H-NMR	Untargeted	Formate, citrulline, taurine, and isocitrate
Perng et al. 2019, Mexico [39]	242	8–14	FFQ	Fasting serum	LC/MS	Untargeted	Girls: 5-methyl-tetrohydrofolate, phenylephrine, urate, nonanoate, deoxyuridine, and sn-glycero-3-phosphocholine Boys: 2-piperidinone, octanoylcarnitine, and catechol
Logue et al. 2020, NL [36]	79	19–70	7-d DR	24-h urine	LC-MS	Targeted	acesulfame-K, saccharin, cyclamate, and sucralose steviol glycosides
Logue et al. 2017, NL [37]	21	25.7 (4.9) [a]	7-d DR	Fasting spot and 24-h urine	LC-MS	Targeted	Acesulfame-K, saccharin, sucralose, cyclamate, and steviol glycosides
Sylvetsky et al. 2017, USA [43]	18	18–35	7-d DR	Spot urine	LC/MS	Targeted	Sucralose

[a] and [b]—values are mean (standard deviation) and median (interquartile range), respectively; $\delta^{13}C$—carbon isotope ratio biomarker, ^{13}C:^{12}C; AAs —amino acids; IRMS—isotope ratio mass spectrometry; GC-IRMS—gas chromatography with IRMS; NA-SIMS —natural abundance stable isotope mass spectrometry; H-NMR—proton (hydrogen) nuclear magnetic resonance; LC/MS—liquid chromatography–mass spectrometry; NL—Netherlands; 24-h recall—24 h dietary recall records; 4-d/7-d DR—4/7 day dietary records; FFQ—food frequency questionnaires. ** This study was conducted in postmenopausal women and reported negative results that, $\delta^{13}C$ was not associated with an intake of sugar, both total and AS/SSBs.

3.1. Carbon Isotope Based Biomarkers for SSBs Intake

The stable carbon isotope ratio ^{13}C:^{12}C, denoted as $\delta^{13}C$ values in blood samples, plasma glucose, hair, and alanine, is significantly correlated with SSB intake [29–32,34,35,38,40,42,44,45]. Even though these studies were carried out in different settings and populations, they employed a targeted approach for biomarker discovery. Davy et al. [30] investigated the $\delta^{13}C$ values of non-fasting fingerstick blood, complemented by four-day food intake records, in healthy participants. To minimize the order effects due to the sequence of dietary intake assessment, investigators randomly assigned participants to either of the two sequences for their laboratory visits, which determined the sequence of the beverage intake assessment and the four-day food intake assessment. Habitual intake of SSBs in the past month was assessed with a separate questionnaire. The $\delta^{13}C$ values were correlated with the AS and SSB intakes [30].

Davy et al. [31], in a separate study, investigated the $\delta^{13}C$ values of fasting fingerstick blood in a randomized controlled trial, using predominantly obese participants to assess whether a 6-month intervention for reducing SSBs intake was reflected on the $\delta^{13}C$ values. This study concluded that, indeed, changes in $\delta^{13}C$ values were associated with the AS and SSBs intake, supporting $\delta^{13}C$ as an objective biomarker of AS and SSBs intake. Similarly, Fakhouri et al. [32] examined the $\delta^{13}C$ values of the serum in response to an 18-month behavioral intervention program for reducing the SSBs intake in adults. Analyses of blinded serum samples confirmed the mean change in $\delta^{13}C$ values, consistent with the self-reported dietary intake SSBs—further confirming earlier studies that $\delta^{13}C$ values could be used to measure small changes in the intake of AS or SSBs.

Nash et al. [45] compared the dietary intake of sugars (total, added, and SSBs) among the Yup'ik people, as reflected in the $\delta^{13}C$ values in the red blood cells, hair, and fasting plasma glucose. Their dual-isotope model approach measured the values of both the $\delta^{13}C$ and stable isotopes of nitrogen, $\delta^{15}N$, which controlled feeding studies have suggested as a potential biomarker for the dietary intake of meat and fish [40,41]. As other dietary components such as animal protein, honey, and beet sugar may confound the association between the AS/SSBs intake and $\delta^{13}C$ values [23], there is potential utility of the dual-isotope method. Moreover, given that the $\delta^{13}C$ and $\delta^{15}N$ values in the red blood cells, serum, and hair have been shown to be correlated [46,47], Nash et al. [45] further examined whether the dual-isotopic model explained a similar variance in the intake of total sugars, AS, and SSBs, as reflected in the red blood cells, plasma, and hair. They observed that the three models using red blood cells, plasma, or hair isotopes explained nearly similar amounts of variance in the dietary intake of total sugar, AS, and SSBs. The strongest associations of sugar intake and $\delta^{13}C$ values were observed in red blood cells and hair samples. There were strong, positive correlations in $\delta^{13}C$ and $\delta^{15}N$ values of red blood cells, plasma, and hair. Collectively, these results demonstrated that the $\delta^{13}C$ biomarker, as reflected in red blood cells, plasma, and hair, but not in the fasting plasma glucose, may be useful in assessing the sugar intake in this Alaska Native community.

Votruba et al. [40] used the dual-isotope model approach to measure the values of $\delta^{13}C$ and $\delta^{15}N$ in the red blood cells, plasma, and hair as potential biomarkers for the dietary intake of SSBs, fish, and meat in a 12-week controlled feeding trial. In this study, they observed that the $\delta^{13}C$ values were significantly elevated by the dietary intake of SSBs and meat, while the $\delta^{15}N$ values were significantly associated with the dietary intake of fish and meat. Specifically, the plasma $\delta^{15}N$ predicted the dietary intake of fish (area under the receiver operating curve (AUC) = 0.97) and meat (AUC = 0.92), while plasma $\delta^{13}C$ predicted the SSBs intake (AUC = 0.78). In all of the sample types—red blood cells, plasma, and hair—the dual-isotope approach accurately distinguished consumers of meat and fish, with a modest discrimination power for consumers of SSBs [40].

Two studies measured the $\delta^{13}C$ of alanine as a potential biomarker of SSB intake. Choy et al. [29] investigated the association between the $\delta^{13}C$ of nonessential amino acids ($\delta^{13}C$ NEAA) in red blood cells and the intake of total sugar, AS, and SSBs, as well as the foods rich in animal protein such as corn-fed meats, fish, and marine mammals. Of the

non-essential amino acids considered (alanine, aspartate, glutamate, glycine, proline, and serine), only the $\delta^{13}C$ of alanine was strongly associated with sugar intake—total sugar, AS, and SSBs—with a moderate association between the $\delta^{13}C$ of proline and SSB intake only [29]. In a subset of the study population (n = 30), $\delta^{13}C$ of alanine in red blood cells was correlated with $\delta^{13}C$ of alanine in hair samples, and intake of SSBs. As the intake of meat and fish may also elevate $\delta^{13}C$ values in some populations [23,48], Choy et al. [29] further tested the specificity of the $\delta^{13}C$ of alanine for SSB intake by modelling the $\delta^{13}C$ of alanine as a dependent variable and adding SSBs; commercial meat, fish, and marine mammals; and intake of corn as independent variables. Unlike the findings of Votruba et al. [40], the $\delta^{13}C$ values of alanine were significantly associated with SSB intake only, but not with any other dietary component, including meat and fish [29]. These findings were replicated in a recent two-week controlled feeding trial in postmenopausal women [42]. The $\delta^{13}C$ of seven amino acids (alanine, glycine, valine, leucine, isoleucine, proline, and phenylalanine) in the fasting serum were measured. Like in Choy et al. [29], AS intake was associated with elevated values of $\delta^{13}C$ of alanine, but was not associated with meat or any other animal protein. These two controlled studies demonstrated the specificity of $\delta^{13}C$ of alanine to AS.

Hedrick et al. [34] compared $\delta^{13}C$ values of fingerstick blood with self-reported AS and SSBs intake in a cross-sectional study of adults who consumed at least 200 kcal/d from SSBs. In their multiple linear regression of $\delta^{13}C$ values on other variables, they observed a significant variation in $\delta^{13}C$ values across different age groups, indicating the highest intake of SSBs and AS in younger adults. Overall, SSB intake was significantly associated with $\delta^{13}C$ values. Similarly, MacDougall et al. [38] explored the comparative validity, reliability, and sensitivity of $\delta^{13}C$ values to reflect AS and SSBs intake in children and adolescents over a 3-week period. Their findings confirmed that $\delta^{13}C$ values discriminated between high and low consumers of SSBs and between high and low consumers of AS in general (AUC = 0.75 and AUC = 0.62, respectively). In a similar study in adolescents by Liu et al. [35], but using a controlled feeding design, the $\delta^{13}C$ values of the fasting fingerstick blood reflected changes in AS and ASSB intakes in different feeding periods. Valenzuela et al. [44] also focused on adolescents, and measured multiple stable isotopes, namely $\delta^{13}C$, and $\delta^{15}N$, and stable isotopes of sulfur $\delta^{34}S$, in hair and breath samples in order to evaluate the potential biomarkers for protein and carbohydrate dietary components. In this study, the intake of SSBs and C4 derived-sweets was associated with $\delta^{13}C$ values from the carbon dioxide in the breath samples, both in the baseline (morning upon waking up) and post-lunch samples (1–2 h after lunch), showing the strongest correlations in the baseline samples [44]. Expectedly, significantly elevated $\delta^{13}C$ values were observed among Hispanic children who were also reported to have a higher consumption of SSBs relative to non-Hispanic white children [44]. Additionally, the $\delta^{13}C$ values in the hair samples were also significantly correlated with the baseline breath samples.

Yun et al. [41] examined whole serum in postmenopausal women in a 2-week controlled diet study. This was the only study that found no association between sugar intake and $\delta^{13}C$ values. Yun et al. [41] measured the values of multiple isotopes, $\delta^{13}C$, $\delta^{15}N$, and $\delta^{34}S$, in relation to the habitual intake of total sugars, AS, SSBs, animal protein, fish/seafood, red meat, dairy, poultry, and eggs. While $\delta^{15}N$ predicted the intake of fish/seafood, $\delta^{13}C$ moderately predicted the intake of red meat and eggs, but did not meet the biomarker threshold for the intake of sugars—total, AS, and SSBs [41]. It should be emphasized that the population in this study had limited heterogeneity in their diet, as some dietary components such as AS and SSBs were consumed in low amounts.

3.2. Other Candidate Biomarkers of SSBs Intake

Some studies used untargeted metabolomics approaches to discover panels of metabolites in biosamples that could indicate the dietary intake of SSBs. Gibbons et al. [33] identified a panel of four metabolites (i.e., formate, citrulline, taurine, and isocitrate) that were significantly associated with SSB intake. They further validated these metabolites in a small acute intervention study using first-void-urine and postprandial urine samples

collected at time intervals of 2, 4, and 6 h after SSB intake (i.e., a 330 mL of caloric cola). Elevated levels of the four biomarkers were recorded in the urine samples following the acute consumption of sweetened cola, and their presence was further confirmed in the chemical analysis of the cola drink [33].

Perng et al. [39], using an untargeted approach, also identified a novel set of metabolites associated with the intake of SSBs using fasting serum samples and an FFQ instrument for the intake assessment in children and adolescents. In this study, SSBs included non-diet sodas, fruit juices with AS, and any other beverage (e.g., tea, coffee, or water) with AS. The authors discovered sex-specific panels of biomarkers that were associated with SSB intake. They reported six biomarkers in girls—5-methyl-tetrohydrofolate, phenylephrine, urate, nonanoate, deoxyuridine, and sn-glycero-3-phosphocholine—and three biomarkers in boys—2-piperidinone, octanoylcarnitine, and catechol.

3.3. Candidate Biomarkers of LCSBs Intake

Three studies investigated the potential biomarkers of low-calorie sweeteners (LCSs) commonly used in LCSBs, identifying urinary excretion of acesulfame-K, saccharin, cyclamate, sucralose, and steviol glycosides among the consumers of LCSs/LCSBs [36,37,43]. Logue et al. [37] investigated the urinary excretion of commonly used LCSs following dietary exposure to LCSBs, using a double-blind, randomized crossover dose−response study. For method development and validation, participants (n = 12) were advised verbally and through written materials to avoid the intake of foods and beverages known to contain the five LCS compounds, at least 3 days before the 24-h urine protocol date. After the analyses, samples without concentrations of LCS (n = 6) were adopted for method validation. For the dose−response study, 21 participants were examined in a double-blind, randomized crossover design, lasting 3 weeks, during which participants consumed three doses of five LCSs, namely acesulfame-K, saccharin, sucralose, cyclamate, and steviol glucuronide [37]. Fasting spot and 24-h urine samples were collected at each dosing date. The 500 mL LCSBs were consumed over two consecutive days at specific times during the study period, but for the purpose of blinding the participants, 75 mL of an orange Cordial was added during LCSBs preparation. As long as the consumption did not exceed 500 mL within the 24-h period, participants were encouraged to assume normal patterns of beverage intake throughout the day. Regression analyses with the LCS dose set as the dependent variable and 24-h urinary concentrations of the LCS compounds as the independent variable explained 99% variability for acesulfame-K, 87% for saccharin, 35% for sucralose, 91% for cyclamate, and 75% steviol glucuronide [37]. These compounds were indicative of LCSBs intake.

In a separate study, Logue et al. [36] further investigated the use of a 24-h urinary biomarker approach to detect dietary exposure to LCSB in two adult population-based studies, targeting the five LCSs investigated previously in their controlled study [37]. The 24-h urinary biomarker was compared with LCSB consumption, as self-reported in 7-d food diaries of the participants (n = 79), who were randomly selected from a large study regarding the prevalence of the widespread consumption of LCSs (n = 357). Participants were grouped into consumers and non-consumers of LCSBs on the urine protocol date. The novel urinary biomarker approach identified proportions of consumers of LCSBs enriched with various sweeteners, namely saccharin (82%), acesulfame-K (51%), cyclamates (34%), sucralose (30%), and steviol glycosides (11%) [36].

Sylvetsky et al. [43] investigated whether non-consumers of LCSs could be correctly characterized as unexposed using the urinary biomarker approach, in a small randomized controlled trial lasting two weeks. Participants were scheduled to attend three visits—all of which were one week apart for urine sample collection and other measurements. As they were confirmed as non-consumers of LCSs during recruitment into the study, participants were counselled to avoid dietary intake of LCSs. At baseline, their dietary intake was also recorded. After a 1-week run-in period, using sex-matched paired design, participants were randomly assigned to consume diet soda containing sucralose or unsweetened carbonated

water, three times a day for a week. Other dietary components were also reviewed if they contained sucralose. At the end of the trial period, the urinary sucralose concentrations in the exposed group were consistent with the LCSB dietary intake–significantly higher than the expected residual sucralose from the occasional consumption of other dietary components containing sucralose [43].

3.4. Evaluation of Validity of Candidate Biomarkers

Table 2 summarizes the results of the evaluation of the candidate biomarkers for the dietary intake of SSBs. The number of times a compound is rated "Y" across validation criteria reflects the current validity of the candidate biomarker, while the "N" and "U" ratings represent areas where more research should be conducted. Candidate biomarkers $\delta^{13}C$ and $\delta^{13}C$ of alanine had the highest validity, with an affirmative rating on the specificity, dose−response, time−response, robustness, reliability, stability, and analytical performance. This carbon isotope ratio biomarker was also studied in many studies, consistently reporting an association with the dietary intake of SSBs or AS [29–32,34,35,38,40,42,44,45].

Evidence of the $\delta^{13}C$ of alanine as a potential biomarker for SSBs [29,42,49] is also consistent with the long established glucose−alanine cycle in humans. The glucose−alanine cycle explains the link between carbohydrate and amino acid metabolism, in which alanine is synthesized from pyruvate, a product of glycolysis. The biochemical plausibility of $\delta^{13}C$ of alanine is, therefore, demonstrable. The C4 derived AS has distinctly high $\delta^{13}C$ values compared with any other dietary source, which proves the distal cause of the biomarker signal, while the proximal link between serum alanine and glucose is explained by the glucose−alanine cycle [50]. This also improves our understanding on the accumulating evidence demonstrating strong $\delta^{13}C$ of alanine correlation with dietary AS, but not with other dietary components [41,42,49].

Uncertainty on the validity of the $\delta^{13}C$ and $\delta^{13}C$ of alanine as biomarkers of SSB intake remains regarding reproducibility across laboratories, as inter-laboratory results have not been described in literature. To fulfil this validation criterion, targeted analysis of the candidate biomarker in common set of samples is recommended, maintaining blind testing across testing laboratories [26]. If an untargeted metabolomics approach is used, a standardized analytical approach should be used by all participating laboratories.

Even though formate, citrulline, taurine and isocitrate, were discovered in an observational study and were validated in a small intervention study [33], these candidate biomarkers were rated 4/9, meeting criteria for dose−response, single meal time−response, robustness, and analytical performance. Lastly, 5-methyl-tetrohydrofolate, phenylephrine, urate, nonanoate, deoxyuridine, sn-glycero-3-phosphocholine, 2-piperidinone, octanoylcarnitine, and catechol showed the lowest validity scores, with a positive rating on robustness and analytical performance only. The low scores identified areas of further research to improve the validity of these candidate biomarkers for the dietary intake of SSBs.

A summary of the evaluation of the validity of the candidate biomarkers for the dietary intake of LCSBs is provided in Table 3. Briefly, acesulfame-K, saccharin, sucralose, cyclamate, and steviol glucuronide showed moderate validity (6/9) in predicting LCSBs intake. All of these compounds are commercially used as low-calories sweeteners. As such, their plausibility as biomarkers of specific LCSBs is fulfilled, but additional qualitative assessments of the dietary intake should rule out other dietary sources. Uncertainty remains regarding their kinetics after repeated or habitual intake, as the compounds were assessed in urine, which reflects recent intake. Accumulation of the compounds as a consequence of habitual intake is inconclusive, as none of these studies investigated the usual intake. Moreover, evidence on the stability and reproducibility of these compounds in the same set of samples across various laboratories has not been described.

Table 2. Evaluation of the validity of the identified candidate biomarkers for dietary intake of SSBs.

Compound/Metabolite	HMDB ID	Sample Type	Validation Criteria										
			1	2	3a	3b	4	5	6	7	8	Max. Points = 9	References
$\delta^{13}C$	-	RBCs, plasma, breath, hair	Y	Y	Y	Y	Y	Y	Y	Y	U	8	[6,23,30–32,34,35,38,40,41,44,45,48,51]
$\delta^{13}C$ of alanine	HMDB0000161	Blood, serum, hair	Y	Y	Y	Y	Y	Y	Y	Y	U	8	[29,42,49,50,52]
Formate	HMDB0000142	Urine	N	Y	Y	U	Y	U	U	Y	N	4	[33,53]
Citrulline	HMDB0000904	Urine	N	Y	Y	U	Y	U	U	Y	N	4	[33,54]
Taurine	HMDB0000251	Urine	N	Y	Y	U	Y	U	U	Y	N	4	[33,55]
Isocitrate	HMDB0000193	Urine	N	Y	Y	U	Y	U	U	Y	N	4	[33,56]
5-Methyl-tetrohydrofolate	HMDB0001396	Serum	N	Y	U	U	Y	U	U	Y	N	2	[39]
Phenylephrine	HMDB0002182	Serum	N	U	U	U	Y	U	U	Y	N	2	[39]
Urate	HMDB0000289	Serum	N	U	U	U	Y	U	U	Y	N	2	[39]
Nonanoate	HMDB0031264	Serum	N	U	U	U	Y	U	U	Y	N	2	[39]
Deoxyuridine	HMDB0000012	Serum	N	U	U	U	Y	U	U	Y	N	2	[39]
Sn-glycero-3-phosphocholine	HMDB0000086	Serum	N	U	U	U	Y	U	U	Y	N	2	[39]
2-Piperidinone	HMDB0011749	Serum	N	U	U	U	Y	U	U	Y	N	2	[39]
Cctanoylcarnitine	HMDB0000791	Serum	N	U	U	U	Y	U	U	Y	N	2	[39]
Catechol	HMDB0240490	serum	N	U	U	U	Y	U	U	Y	N	2	[39]

SSBs—sugar-sweetened beverages; RBCs – red blood cells; Y—yes; N—no; U—unknown/uncertain (validation criteria adapted from [26]). 1: Plausibility—Is the marker compound plausible as a specific BFI for the food or food group (chemical/biological plausibility)? 2: Dose Response—Is there a dose−response relationship at the relevant intake levels of the targeted food (quantitative aspect)? 3: Time Response—(a) Single dose: meal time−response relationship of the BFI has been described for a defined sample type and time window in a meal study. (b) Multiple doses: the kinetics of the BFI after repeated intakes has been described for a defined sample type in a meal study or the accumulation of BFI in certain sample types has been observed. Is the biomarker kinetics for the repeated intake of the food/food group described adequately providing the frequency of sampling needed to assess the habitual intake (e.g., cumulative aspects). 4: Robustness—Has the marker been shown to be robust after the intake of complex meals reflecting the dietary habits of the targeted population? 5: Reliability—Has the marker been shown to compare well with other markers or questionnaire data for the same food/food group (reliability)? 6: Stability—Is the marker chemically and biologically stable during bio specimen collection and storage, making measurements reliable and feasible? 7: Analytical Performance—Are analytical variability, accuracy, sensitivity, and specificity known to be adequate for at least one reported analytical method? 8: Reproducibility—Has the analysis been successfully reproduced in another laboratory?

Table 3. Evaluation of the validity of the identified candidate biomarkers for the dietary intake of LCSBs.

Compound/Metabolite	HMDB ID	Sample Type	Validation Criteria										
			1	2	3a	3b	4	5	6	7	8	Max. Points = 9	References
Acesulfame-K	HMDB0033585	Urine	Y	Y	Y	U	Y	Y	U	Y	U	6	[36,57]
Saccharin	HMDB0029723	Urine	Y	Y	Y	U	Y	Y	U	Y	U	6	[36,37,58]
Cyclamate	HMDB0031340	Urine	Y	Y	Y	U	Y	Y	U	Y	U	6	[36,37,57,59,60]
Sucralose	HMDB0031554	Urine	Y	Y	Y	U	Y	Y	U	Y	U	6	[36,43,57,61]
Steviol glycosides	HMDB0036707	Urine	Y	Y	Y	U	Y	Y	U	Y	U	6	[14,37,57,62]

LCSBs—low-calorie sweetened beverages; Response: Y—yes; N—no; U—unknown/uncertain (validation criteria adapted from [26], as explained under Table 2).

3.5. Risk of Bias and Quality of Study Assessment

The risk of bias and quality of evidence assessment for the included studies is presented in Supplementary Table S1 for the observational studies, and Supplementary Table S2 for the controlled intervention studies. Overall, the quality assessment scores for the observational studies ranged between 9 and 12.5 out of the attainable 14.5 points for this study design. Therefore, they were rated moderate to high quality. The quality scores for the interventional studies ranged between 7.5 and 9.5 out of the attainable 13 points for the controlled intervention studies. Given the high threshold for assessing the risk of bias and outcomes in controlled studies, evidence from these interventional studies was considered to be of a moderate quality.

4. Discussion

The main ingredient of SSBs is AS, and it is estimated that nearly half of AS is consumed through SSBs [63]. Long established biomarkers for sugar intake are 24-h urinary sucrose/fructose biomarkers [64–67] or sucrose/fructose in spot urine [68,69]. The sucrose/fructose biomarker, however, reflects the total sugar intake from all dietary components; it lacks specificity for AS and is thus not plausible for assessing SSBs intake. The carbon isotope method, demonstrated by elevated carbon isotope signatures, e.g., in urine, serum amino acids, red blood cells, or hair, reflects the dietary intake of AS [29–32,34,35,38,40,42,45]. When SSBs, which are highly correlated with AS, are consumed, the carbon isotopes are also absorbed and become available in the tissues. The $\delta^{13}C$ biomarker values reflect the carbon isotopic composition of the plant from which the AS was refined, which could either be C3 or C4 photosynthetic plants [23]. For some regions, an illustrative example being the USA, sweeteners are mostly refined from corn syrup and cane sugar, which all utilize the C4 photosynthetic pathway [23]. SSBs with AS derived from C4 plants have high C4 isotope signatures. We consider this is the reason that all studies included in this review on the $\delta^{13}C$ biomarker for SSBs are based on USA populations. In regions where the main source of AS is sugar beets (e.g., in Europe), which utilize C3 photosynthetic pathway, $\delta^{13}C$ is not an appropriate biomarker for AS intake [6,35,50,70]. There are differential biochemical processes in C4 and C3 plants, in which C4 plants extract heavier $^{13}CO_2$ from the atmosphere than C3 plants. Sugars refined from C4 plants are consequently more enriched with ^{13}C isotopes. This means that the stable carbon isotope method can be applied to predict high consumers of SSBs—containing C4 derived AS—because of their elevated $\delta^{13}C$ values [23,70].

The use of a stable carbon isotope as a biomarker of SSB intake has specific strengths. The fingerstick sample collection method is simple to conduct a minimally invasive and not burdensome task for the participants [23,70,71]. The 24-h urinary collection may be burdensome for some participants, eliciting concerns about compliance. Stable carbon isotopes of hair and breath, as shown in [44], are especially useful in large-scale epidemiologic studies. As the carbon isotopes integrate diet over a long period, typically weeks to months [40,45,72], they provide better estimates of habitual sugar intake compared with

fructose and sucrose urinary biomarkers, which integrate short-term dietary intake. $\delta^{13}C$ has also been shown to be stable and readily assayed in tissues such as red blood cells, serum amino acids, and human hair, for short-term and long-term exposure to AS [23,40]. As $\delta^{13}C$ is more enriched in C4 photosynthetic plants compared with C3 plants, it can discriminate between AS and naturally occurring sugars (e.g., from fruits), which are mostly C3 plants. As such, beverages with high fruit concentrates are shown to have a significantly lower $\delta^{13}C$ content than beverages enriched by cane sugar or corn syrup [23].

In a recent controlled feeding study [42], the biological plausibility for use of $\delta^{13}C$ alanine as a biomarker for AS was demonstrated, as amino acid carbon isotope signatures discriminated AS from red meat/protein intake; specifically, the $\delta^{13}C$ of alanine reflected a primary intake of AS. Additionally, the rest of the amino acids carbon isotope values showed an inverse association with sugar intake—total sugar, AS and SSBs—but a positive association with the intake of animal proteins and animal-derived dietary components such as red meat [42]. These findings are consistent with the results of Choy et al. [29], which demonstrated the $\delta^{13}C$ of alanine of red blood cells was significantly associated with total sugar intake, AS, and SSBs, notwithstanding their differences in analytical approach, population, and dietary assessment methods. By targeting specific serum amino acids only, Yun et al. [42] further advanced the field, as previous approaches based on whole serum suggested that $\delta^{13}C$ values were also associated with other dietary factors such as animal proteins sources, e.g., meat and other protein intake [41].

A recent study [49], not included in Table 1 because it was published outside the records search period, corroborates evidence on the specificity of $\delta^{13}C$ alanine for AS and SBBs. In this study, the $\delta^{13}C$ of alanine and $\delta^{13}C$ of glutamate were individual predictors of SSBs intake, with a predictive accuracy of AUC $\geq$ 0.97 and no evident association with meat intake. The findings also suggested that using a multiple amino acids approach could improve the biomarker estimation of the SSBs intake [49]. On the other hand, the $\delta^{13}C$ of essential amino acids, especially the $\delta^{13}C$ of leucine, was the most promising predictor of meat intake (AUC $\geq$ 0.92). Moreover, an important addition of this study to the current literature was the observation that the concentration of $\delta^{13}C$ of non-essential amino acids is not influenced by meat intake, reflected greater sensitivity, and was more specific to SSBs intake, unlike when the $\delta^{13}C$ values of the total tissue (plasma and red blood cells) were measured [49]. Previous studies showed that $\delta^{13}C$ total tissue was more strongly related to meat and/or animal protein intake than AS and SSBs [40,41]. The results of Yun et al., Choy et al., and Johnson et al. collectively validate the specificity of $\delta^{13}C$ alanine as a biomarker for SSBs intake and not animal proteins [29,42,49]. Given that these three studies were conducted in diverse populations, this also demonstrates the robustness of this biomarker. What remains inconclusive from these studies is whether individual or multiple amino acid $\delta^{13}C$ values best estimate AS and SSB intake, given that they used different blood fractions, derivatization, and analyses of amino acids, leading to slightly different sets of amino acids that were reliably measured [49].

A major limitation of the $\delta^{13}C$ biomarker is its limited specificity and sensitivity with respect to AS and metabolically different sources of such sweeteners [6,38,50]. Theoretically, $\delta^{13}C$ values reflect all dietary items from plants utilizing the C4 photosynthetic pathway. Hence, the biomarker may not reflect the SSBs intake alone. Moreover, the $\delta^{13}C$ is limited to AS refined from C4 plants (cane sugar and corn syrup), and not sugar refined from C3 plants like the beet sugar [6,23]. Therefore, the application of the $\delta^{13}C$ biomarker of SSBs is limited to populations that consume sugars refined from C4 sources. Furthermore, even though dietary glucose and fructose moieties are assumed to have similar metabolic fates, this is unpredictable and unlikely to hold true if high inter-individual variability exists [6]. Additionally, none of the studies included in this review demonstrated the validity of the $\delta^{13}C$ biomarker in populations that consume a large proportion of dietary energy from corn-based foods. The values of $\delta^{13}C$ in the blood samples may also be influenced by the dietary intake of meat from livestock fed corn-based diets, which potentially confounds the specificity of the $\delta^{13}C$ biomarker [6,23,50]. Attempts to control this potential confounder

with use of the nitrogen isotope, $\delta^{15}N$—found in proteins and not in sugar—yielded mixed findings [6,48,51], with nearly two thirds variation in the self-reported dietary intake of AS being attributed to other factors beyond the scope of the $\delta^{13}C$ biomarker analysis approach [48]. The use of $\delta^{15}N$ in another study only marginally increased the correlation between AS and $\delta^{13}C$ values [51]. Given that the majority of studies relied on self-reported dietary data, this warrants further analyses in controlled feeding settings. However, it should be emphasized that in a recent controlled feeding study, $\delta^{13}C$ in the serum amino acids rather than in whole serum or in red blood cells, was correlated with AS intake but not dietary intake of animal protein or red meat [42,49].

Even though the $\delta^{13}C$ alanine biomarker for SSBs, as proposed by Choy et al. [29], is biochemically plausible and specific as validated in controlled feeding studies [40,42,49], values of the $\delta^{13}C$ of alanine may be influenced by complex metabolic processes along the chain of inference, including extraneous factors such as fasting state, dietary composition, overweight, and obesity [50]. For example, the proximal link between $\delta^{13}C$ of glucose and serum alanine in the glucose−alanine cycle. Additionally, because of the lack of accepted reference methods for estimating the habitual AS intake, validation studies rely on short-term controlled feeding measures, as observed in the Yun et al. study [42]. For example, they conducted a controlled feeding study for 2 weeks, yet the half-life of $\delta^{13}C$ in plasma is estimated at 2.5 weeks [40]. Hence the dietary period falls short of the residence time of the serum $\delta^{13}C$ of alanine [40,42]. This potentially biased the AS-$\delta^{13}C$ association towards the null by attenuating the effect sizes [49]. In another study, it was determined that stable isotope ratio signatures in the plasma and red blood cells required 8–12 weeks and 15–19 weeks, respectively, to reach isotopic equilibration [40]. In the study of Johnson et al. [49], the carbon isotope ratios of the amino acids in the red blood cells were not at or near equilibrium at the end of the 12-week study. Therefore, the process of validating stable isotope biomarkers using short-term controlled feeding programs raises methodological concerns [50].

As for the panel of biomarkers identified in Gibbons et al. [33], none of these candidate biomarkers have been validated by another study; thus, more mechanistic investigations, besides the validation process, are warranted. Their presence in urine could be confounded by extraneous factors other than the intake of SSBs. The proposed compounds are not normally added in their pure form during the processing of cola drinks [73]. For instance, formate has been cited as an intermediate in normal metabolism, produced from different metabolic sources [53]. Taurine, commonly used as a dietary supplement in energy drinks, is also present in other food items, e.g., naturally occurring in shellfish, meat, and dairy products [55], which limits its specificity for AS. Similarly, watermelons are known to be rich dietary sources of citrulline [54]. Isocitrate, which essentially is an isomerized citrate, is used as a food additive, but dietary sources includes fruit juices, especially blackberries and vegetables such as carrots [56]. This panel of biomarkers, therefore, requires more investigations regarding their biological plausibility and robustness in other study settings. Importantly, validation study designs should account for the potential confounding effect of other dietary sources, as well as intermediates of metabolic processes that may be transformed into these candidate biomarkers. Similarly, this applies to the set of metabolites indicative of SSBs consumption in the study of Perng et al. [39]. Some of the candidate biomarkers (e.g., nonanoate) are dietary supplements and may be derived from other food groups, including fruit flavored SSBs and alcoholic drinks [39]. As none of them has been validated in intervention studies or any other general population study, their specificity and sensitivity for AS or SSBs, therefore, remains inconclusive.

Until recently, there were almost non-existent metabolomics, population-based studies on the biomarkers of dietary LCSBs [73]. The present review identified three recent studies that have explored this research area, identifying common LCSs namely, acesulfame-K, saccharin, cyclamate, sucralose, and steviol glycosides in urine as indicative of LCSB intake. These findings support the hypothesis that a biomarker approach has potential to objectively assess the intake of common LCSBs, especially, given that most of these LCSs

are excreted unchanged in urine [73]. Moreover, these compounds (i.e., acesulfame-K, saccharin, sucralose, cyclamate, and steviol glycosides) are not produced endogenously, and are highly specific to the ingestion of the parent compound [73]. However, as these LCSs are also used in other foods as sweeteners, relying on LCSBs alone as the surrogates for LCS intake may be misleading, because this biomarker approach does not discriminate specific sources of LCSs within a diet. More comprehensive methods are needed for the assessment of dietary intakes, including qualitative data and review of all foods for presence of LCSs [36]. As observed in Sylvetsky et al. [43], the presence of sucralose in urine of LCSs non-consumers confirms that people consume LCSs inadvertently in other dietary sources other than LCSBs. Other non-dietary sources of LCSs, such as personal care products (e.g., oral hygiene products), may also potentially bias the results [43,57]. Taken together, even though the urinary excretion of LCSs reflects its dietary intake [36,37], this novel approach should be further developed to account for inter-and intra-individual variations with respect to dietary intake and urinary excretions in different study settings, populations, and health status [36].

5. Conclusions

This review observed that the most promising candidate biomarker of SSBs is $\delta^{13}C$, with $\delta^{13}C$ of alanine being the most robust, sensitive, and specific to SSBs. Improved estimation of the SSB intake may be realized by measuring the $\delta^{13}C$ of multiple non-essential amino acids. Stable carbon isotopes in the total tissues, such as plasma and red blood cells, were observed to be confounded by other dietary components, particularly, meat, fish, and/or animal protein, therefore, showed modest discrimination power for AS and SSBs intake. A major limitation in the application of carbon isotope-based biomarkers is the inability to detect AS refined from sources that utilize the C3 photosynthetic pathway and other sources. The panel of candidate biomarkers of SSBs, as identified via untargeted metabolomics studies, require further investigation regarding their biochemical plausibility and validation in dose−response studies before they can be used in epidemiological studies. We also observed that LCSs, particularly acesulfame-K, saccharin, sucralose, cyclamate, and steviol glucuronide, may predict the intake of LCSBs in regions where such sweeteners are approved for commercial use. This is a promising area of research, as some of LCSs compounds are excreted unchanged via urine, are not produced endogenously in other metabolic processes, and are highly specific to dietary intake. However, other sweeteners may undergo metabolism into metabolites chemically indistinguishable from those produced from other dietary sources. The differences in the metabolic fates of LCSs should, therefore, be considered in biomarker discovery studies. Moreover, these sweeteners are also used in other foods. As such, the urinary concentration of these metabolites may not reflect the LCSB intake alone, unless qualitative data on other food group intake are properly assessed. In addition, given that these are urinary-based biomarkers that reflect short-term exposures, further research needs to characterize the habitual intake of LCSBs.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/metabo11080546/s1. Table S1: Combined quality assessment using NUTRIGRADE and BIOCROSS tools for evaluation of observational studies and biomarkers in human research, Table S2: Combined quality assessment using NUTRIGRADE and BIOCROSS tools for evaluation of controlled intervention studies and biomarkers in human research.

Author Contributions: Conceptualization, S.M., K.O. and U.N.; methodology, S.M. and J.G.; software, S.M. and J.G.; validation, S.M. and J.G.; investigation, S.M. and J.G.; data curation, S.M.; writing—original draft preparation, S.M.; writing—review and editing, S.M., J.G., K.O., A.F., M.S. and U.N.; supervision, K.O. and U.N.; project administration, U.N; funding acquisition, U.N. and A.F. All authors have read and agreed to the published version of the manuscript.

Funding: S.M. is supported by the German Federal Ministry of Education and Research (project number FKZ:01EA1410A) and the Agence Nationale de la Recherche (ANR)/German Research Foundation (DfG) (project: FOOD METabolome as a novel concept to assess dietary exposures in Children, FOODMETCH).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bowman, S.A. Added sugars: Definition and estimation in the USDA Food Patterns Equivalents Databases. *J. Food Compos. Anal.* **2017**, *64*, 64–67. [CrossRef]
2. Guideline: Sugars Intake for Adults and Children. World Health Organization. Geneva. 2015. Available online: https://www.who.int/publications/i/item/9789241549028 (accessed on 25 June 2021).
3. Ernst, J.B.; Arens-Azevedo, U.; Bitzer, B.; Bosy-Westphal, A.; de Zwaan, M.; Egert, S.; Fritsche, A.; Gerlach, S.; Hauner, H.; Heseker, H.; et al. Quantitative recommendation on sugar intake in Germany. *Ernarhrungs Umsch. Int.* **2019**, *66*, M78–M86.
4. Perrar, I.; Schmitting, S.; Della Corte, K.W.; Buyken, A.E.; Alexy, U. Age and time trends in sugar intake among children and adolescents: Results from the DONALD study. *Eur. J. Nutr.* **2019**, *59*, 1043–1054. [CrossRef]
5. Powell, E.S.; Smith-Taillie, L.P.; Popkin, B.M. Added sugars intake across the distribution of US children and adult consumers: 1977–2012. *J. Acad. Nutr. Diet.* **2016**, *116*, 1543–1545. [CrossRef] [PubMed]
6. Louie, J.C. Objective biomarkers for total added sugar intake—Are we on a wild goose chase? *Adv. Nutr.* **2020**, *11*, 1429–1436. [CrossRef] [PubMed]
7. Qin, P.; Li, Q.; Zhao, Y.; Chen, Q.; Sun, X.; Liu, Y.; Li, H.; Wang, T.; Chen, X.; Zhou, Q.; et al. Sugar and artificially sweetened beverages and risk of obesity, type 2 diabetes mellitus, hypertension, and all-cause mortality: A dose–response meta-analysis of prospective cohort studies. *Eur. J. Epidemiol.* **2020**, *35*, 655–671. [CrossRef]
8. Drouin-Chartier, J.-P.; Zheng, Y.; Li, Y.; Malik, V.; Pan, A.; Bhupathiraju, S.N.; Tobias, D.K.; Manson, J.E.; Willett, W.C.; Hu, F.B. Changes in Consumption of Sugary Beverages and Artificially Sweetened Beverages and Subsequent Risk of Type 2 Diabetes: Results from Three Large Prospective U.S. Cohorts of Women and Men. *Diabetes Care* **2019**, *42*, 2181–2189. [CrossRef]
9. Dhingra, R.; Sullivan, L.; Jacques, P.F.; Wang, T.J.; Fox, C.S.; Meigs, J.B.; D'Agostino, R.B.; Gaziano, J.M.; Vasan, R.S. Soft Drink Consumption and Risk of Developing Cardiometabolic Risk Factors and the Metabolic Syndrome in Middle-Aged Adults in the Community. *Circulation* **2007**, *116*, 480–488. [CrossRef]
10. de Koning, L.; Malik, V.S.; Kellogg, M.; Rimm, E.B.; Willett, W.C.; Hu, F.B. Sweetened Beverage Consumption, Incident Coronary Heart Disease, and Biomarkers of Risk in Men. *Circulation* **2012**, *125*, 1735–1741. [CrossRef]
11. Fung, T.T.; Malik, V.; Rexrode, K.; E Manson, J.; Willett, W.C.; Hu, F.B. Sweetened beverage consumption and risk of coronary heart disease in women. *Am. J. Clin. Nutr.* **2009**, *89*, 1037–1042. [CrossRef]
12. Herter-Aeberli, I.; Gerber, P.; Hochuli, M.; Kohler, S.; Haile, S.; Gouni-Berthold, I.; Berthold, H.; A Spinas, G.; Berneis, K. Low to moderate sugar-sweetened beverage consumption impairs glucose and lipid metabolism and promotes inflammation in healthy young men: A randomized controlled trial. *Am. J. Clin. Nutr.* **2011**, *94*, 479–485. [CrossRef] [PubMed]
13. Lin, W.-T.; Kao, Y.-H.; Sothern, M.S.; Seal, D.W.; Lee, C.-H.; Lin, H.-Y.; Chen, T.; Tseng, T.-S. The association between sugar-sweetened beverages intake, body mass index, and inflammation in US adults. *Int. J. Public Health* **2020**, *65*, 45–53. [CrossRef]
14. Pang, M.D.; Goossens, G.H.; Blaak, E.E. The Impact of Artificial Sweeteners on Body Weight Control and Glucose Homeostasis. *Front. Nutr.* **2021**, *7*, 598340. [CrossRef]
15. 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.; et al. Nonnutritive sweeteners and cardiometabolic health: A systematic review and meta-analysis of randomized controlled trials and prospective cohort studies. *Can. Med. Assoc. J.* **2017**, *189*, E929–E939. [CrossRef]
16. Pase, M.P.; Himali, J.J.; Beiser, A.S.; Aparicio, H.J.; Satizabal, C.L.; Vasan, R.S.; Seshadri, S.; Jacques, P.F. Sugar- and Artificially Sweetened Beverages and the Risks of Incident Stroke and Dementia: A Prospective Cohort Study. *Stroke* **2017**, *48*, 1139–1146. [CrossRef]
17. Dalenberg, J.R.; Patel, B.P.; Denis, R.; Veldhuizen, M.; Nakamura, Y.; Vinke, P.C.; Luquet, S.; Small, D.M. Short-Term Consumption of Sucralose with, but Not without, Carbohydrate Impairs Neural and Metabolic Sensitivity to Sugar in Humans. *Cell Metab.* **2020**, *31*, 493–502. [CrossRef]
18. Chazelas, E.; Debras, C.; Srour, B.; Fezeu, L.K.; Julia, C.; Hercberg, S.; Deschasaux, M.; Touvier, M. Sugary drinks, artificially-sweetened beverages, and cardiovascular disease in the nutrinet-sante cohort. *J. Am. Coll Cardiol.* **2020**, *76*, 2175–2177. [CrossRef] [PubMed]
19. Paeratakul, S.; Popkin, B.M.; Kohlmeier, L.; Hertz-Picciotto, I.; Guo, X.; Edwards, L. Measurement error in dietary data: Implications for the epidemiologic study of the diet–disease relationship. *Eur. J. Clin. Nutr.* **1998**, *52*, 722–727. [CrossRef] [PubMed]
20. Krebs-Smith, S.; Graubard, B.; Kahle, L.; Subar, A.; Cleveland, L.; Ballard-Barbash, R. Low energy reporters vs. others: A comparison of reported food intakes. *Eur. J. Clin. Nutr.* **2000**, *54*, 281–287. [CrossRef] [PubMed]
21. Gemming, L.; Ni Mhurchu, C. Dietary under-reporting: What foods and which meals are typically under-reported? *Eur. J. Clin. Nutr.* **2016**, *70*, 640–641. [CrossRef]

22. Guasch-Ferré, M.; Bhupathiraju, S.N.; Hu, F.B. Use of Metabolomics in Improving Assessment of Dietary Intake. *Clin. Chem.* **2018**, *64*, 82–98. [CrossRef]
23. Jahren, A.H.; Bostic, J.N.; Davy, B.M. The potential for a carbon stable isotope biomarker of dietary sugar intake. *J. Anal. At. Spectrom.* **2014**, *29*, 795–816. [CrossRef]
24. Praticò, G.; Gao, Q.; Scalbert, A.; Vergères, G.; Kolehmainen, M.; Manach, C.; Brennan, L.; Pedapati, S.H.; Afman, L.A.; Wishart, D.S.; et al. Guidelines for Biomarker of Food Intake Reviews (BFIRev): How to conduct an extensive literature search for biomarker of food intake discovery. *Genes Nutr.* **2018**, *13*, 1–14. [CrossRef] [PubMed]
25. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. On behalf of the PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med.* **2009**, *6*, e1000097. [CrossRef] [PubMed]
26. Dragsted, L.O.; Gao, Q.; Scalbert, A.; Vergères, G.; Kolehmainen, M.; Manach, C.; Brennan, L.; Afman, L.A.; Wishart, D.S.; Lacueva, C.A.; et al. Validation of biomarkers of food intake—critical assessment of candidate biomarkers. *Genes Nutr.* **2018**, *13*, 14. [CrossRef] [PubMed]
27. Schwingshackl, L.; Knüppel, S.; Schwedhelm, C.; Hoffmann, G.; Missbach, B.; Stelmach-Mardas, M.; Dietrich, S.; Eichelmann, F.; Kontopantelis, E.; Iqbal, K.; et al. Perspective: NutriGrade: A Scoring System to Assess and Judge the Meta-Evidence of Randomized Controlled Trials and Cohort Studies in Nutrition Research. *Adv. Nutr.* **2016**, *7*, 994–1004. [CrossRef]
28. Wirsching, J.; Graßmann, S.; Eichelmann, F.; Harms, L.M.; Schenk, M.; Barth, E.; Berndzen, A.; Olalekan, M.; Sarmini, L.; Zuberer, H.; et al. Development and reliability assessment of a new quality appraisal tool for cross-sectional studies using biomarker data (BIOCROSS). *BMC Med. Res. Methodol.* **2018**, *18*, 122. [CrossRef]
29. Choy, K.; Nash, S.H.; Kristal, A.R.; Hopkins, S.; Boyer, B.B.; O'Brien, D.M. The Carbon Isotope Ratio of Alanine in Red Blood Cells Is a New Candidate Biomarker of Sugar-Sweetened Beverage Intake. *J. Nutr.* **2013**, *143*, 878–884. [CrossRef]
30. Davy, B.M.; Jahren, A.H.; Hedrick, V.E.; Comber, D.L. Association of delta(1)(3)C in fingerstick blood with added-sugar and sugar-sweetened beverage intake. *J. Am. Diet. Assoc.* **2011**, *111*, 874–878. [CrossRef]
31. Davy, B.M.; Jahren, A.H.; Hedrick, V.E.; You, W.; Zoellner, J.M. Influence of an intervention targeting a reduction in sugary beverage intake on the delta13C sugar intake biomarker in a predominantly obese, health-disparate sample. *Public Health Nutr.* **2017**, *20*, 25–29. [CrossRef]
32. Fakhouri, T.H.I.; Jahren, A.H.; Appel, L.J.; Chen, L.; Alavi, R.; Anderson, C.A.M. Serum Carbon Isotope Values Change in Adults in Response to Changes in Sugar-Sweetened Beverage Intake. *J. Nutr.* **2014**, *144*, 902–905. [CrossRef] [PubMed]
33. Gibbons, H.; McNulty, B.; Nugent, A.; Walton, J.; Flynn, A.; Gibney, M.J.; Brennan, L. A metabolomics approach to the identification of biomarkers of sugar-sweetened beverage intake. *Am. J. Clin. Nutr.* **2015**, *101*, 471–477. [CrossRef] [PubMed]
34. Hedrick, V.E.; Davy, B.M.; Wilburn, G.A.; Jahren, A.H.; Zoellner, J.M. Evaluation of a novel biomarker of added sugar intake (delta 13C) compared with self-reported added sugar intake and the Healthy Eating Index-2010 in a community-based, rural U.S. sample. *Public Health Nutr.* **2016**, *19*, 429–436. [CrossRef] [PubMed]
35. Liu, S.V.; Moore, L.B.; Halliday, T.; Jahren, A.H.; Savla, J.; E Hedrick, V.; Marinik, E.L.; Davy, B.M. Short-term changes in added sugar consumption by adolescents reflected in the carbon isotope ratio of fingerstick blood. *Nutr. Health* **2018**, *24*, 251–259. [CrossRef]
36. Logue, C.; Dowey, L.R.C.; Verhagen, H.; Strain, J.J.; O'Mahony, M.; Kapsokefalou, M.; Athanasatou, A.; Gallagher, A.M. A Novel Urinary Biomarker Approach Reveals Widespread Exposure to Multiple Low-Calorie Sweeteners in Adults. *J. Nutr.* **2020**, *150*, 2435–2441. [CrossRef]
37. Logue, C.; Dowey, L.R.C.; Strain, J.J.; Verhagen, H.; McClean, S.; Gallagher, A.M. Application of Liquid Chromatography–Tandem Mass Spectrometry To Determine Urinary Concentrations of Five Commonly Used Low-Calorie Sweeteners: A Novel Biomarker Approach for Assessing Recent Intakes? *J. Agric. Food Chem.* **2017**, *65*, 4516–4525. [CrossRef]
38. MacDougall, C.R.; E Hill, C.; Jahren, A.H.; Savla, J.; Riebl, S.K.; E Hedrick, V.; Raynor, H.; Dunsmore, J.C.; I Frisard, M.; Davy, B.M. The δ13C Value of Fingerstick Blood Is a Valid, Reliable, and Sensitive Biomarker of Sugar-Sweetened Beverage Intake in Children and Adolescents. *J. Nutr.* **2018**, *148*, 147–152. [CrossRef]
39. Perng, W.; Tang, L.; Song, P.X.K.; Goran, M.; Rojo, M.M.T.; Cantoral, A.; Peterson, K.E. Urate and Nonanoate Mark the Relationship between Sugar-Sweetened Beverage Intake and Blood Pressure in Adolescent Girls: A Metabolomics Analysis in the ELEMENT Cohort. *Metabolites* **2019**, *9*, 100. [CrossRef]
40. Votruba, S.B.; A Shaw, P.; Oh, E.J.; A Venti, C.; Bonfiglio, S.; Krakoff, J.; O'Brien, D.M. Associations of plasma, RBCs, and hair carbon and nitrogen isotope ratios with fish, meat, and sugar-sweetened beverage intake in a 12-wk inpatient feeding study. *Am. J. Clin. Nutr.* **2019**, *110*, 1306–1315. [CrossRef]
41. Yun, H.Y.; Lampe, J.W.; Tinker, L.F.; Neuhouser, M.L.; Beresford, S.A.A.; Niles, K.R.; Mossavar-Rahmani, Y.; Snetselaar, L.G.; Van Horn, L.; Prentice, R.L.; et al. Serum Nitrogen and Carbon Stable Isotope Ratios Meet Biomarker Criteria for Fish and Animal Protein Intake in a Controlled Feeding Study of a Women's Health Initiative Cohort. *J. Nutr.* **2018**, *148*, 1931–1937. [CrossRef]
42. Yun, H.Y.; Tinker, L.F.; Neuhouser, M.L.; A Schoeller, D.; Mossavar-Rahmani, Y.; Snetselaar, L.G.; Van Horn, L.V.; Eaton, C.B.; Prentice, R.L.; Lampe, J.W.; et al. The Carbon Isotope Ratios of Serum Amino Acids in Combination with Participant Characteristics can be Used to Estimate Added Sugar Intake in a Controlled Feeding Study of US Postmenopausal Women. *J. Nutr.* **2020**, *150*, 2764–2771. [CrossRef] [PubMed]
43. Sylvetsky, A.C.; Walter, P.J.; Garraffo, H.M.; Robien, K.; I Rother, K. Widespread sucralose exposure in a randomized clinical trial in healthy young adults. *Am. J. Clin. Nutr.* **2017**, *105*, 820–823. [CrossRef] [PubMed]

44. Valenzuela, L.O.; O'Grady, S.P.; Enright, L.E.; Murtaugh, M.; Sweeney, C.; Ehleringer, J.R. Evaluation of childhood nutrition by dietary survey and stable isotope analyses of hair and breath. *Am. J. Hum. Biol.* **2018**, *30*, e23103. [CrossRef] [PubMed]
45. Nash, S.H.; Kristal, A.; Hopkins, S.E.; Boyer, B.B.; O'Brien, D. Stable Isotope Models of Sugar Intake Using Hair, Red Blood Cells, and Plasma, but Not Fasting Plasma Glucose, Predict Sugar Intake in a Yup'ik Study Population. *J. Nutr.* **2014**, *144*, 75–80. [CrossRef]
46. Nash, S.H.; Kristal, A.R.; Boyer, B.B.; King, I.B.; Metzgar, J.S.; O'Brien, D.M. Relation between stable isotope ratios in human red blood cells and hair: Implications for using the nitrogen isotope ratio of hair as a biomarker of eicosapentaenoic acid and docosahexaenoic acid. *Am. J. Clin. Nutr.* **2009**, *90*, 1642–1647. [CrossRef] [PubMed]
47. Kraft, R.A.; Jahren, A.H.; Saudek, C.D. Clinical-scale investigation of stable isotopes in human blood: Delta13C and delta15N from 406 patients at the Johns Hopkins Medical Institutions. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 3683–3692. [CrossRef]
48. Nash, S.H.; Kristal, A.; Bersamin, A.; Hopkins, S.E.; Boyer, B.B.; O'Brien, D. Carbon and Nitrogen Stable Isotope Ratios Predict Intake of Sweeteners in a Yup'ik Study Population. *J. Nutr.* **2013**, *143*, 161–165. [CrossRef] [PubMed]
49. Johnson, J.J.; Shaw, P.A.; Oh, E.J.; Wooller, M.J.; Merriman, S.; Yun, H.Y.; Larsen, T.; Krakoff, J.; Votruba, S.B.; O'Brien, D.M. The carbon isotope ratios of nonessential amino acids identify sugar-sweetened beverage (SSB) consumers in a 12-wk inpatient feeding study of 32 men with varying SSB and meat exposures. *Am. J. Clin. Nutr.* **2021**, *113*, 1256–1264. [CrossRef] [PubMed]
50. O'Connell, T.C. Rough Diamond: A Carbon Isotopic Biomarker of Added Sugar Intake. *J. Nutr.* **2020**, *150*, 2615–2616. [CrossRef]
51. Hedrick, V.E.; Zoellner, J.M.; Jahren, A.H.; Woodford, N.A.; Bostic, J.N.; Davy, B.M. A Dual-Carbon-and-Nitrogen Stable Isotope Ratio Model Is Not Superior to a Single-Carbon Stable Isotope Ratio Model for Predicting Added Sugar Intake in Southwest Virginian Adults. *J. Nutr.* **2015**, *145*, 1362–1369. [CrossRef] [PubMed]
52. Schoeller, D.A. A Novel Carbon Isotope Biomarker for Dietary Sugar. *J. Nutr.* **2013**, *143*, 763–765. [CrossRef] [PubMed]
53. Brosnan, M.E.; Brosnan, J.T. Formate: The Neglected Member of One-Carbon Metabolism. *Annu. Rev. Nutr.* **2016**, *36*, 369–388. [CrossRef] [PubMed]
54. Collins, J.K.; Wu, G.; Perkins-Veazie, P.; Spears, K.; Claypool, P.L.; Baker, R.A.; Clevidence, B.A. Watermelon consumption increases plasma arginine concentrations in adults. *Nutritioon* **2007**, *23*, 261–266. [CrossRef]
55. Wójcik, O.P.; Koenig, K.L.; Zeleniuch-Jacquotte, A.; Costa, M.; Chen, Y. The potential protective effects of taurine on coronary heart disease. *Atherosclerosis* **2010**, *208*, 19–25. [CrossRef]
56. Goldberg, I.; Rokem, J. *Organic and Fatty Acid Production, Microbial*; Elsevier: Amsterdam, The Netherlands, 2009; pp. 421–442.
57. Logue, C.; Dowey, L.C.; Strain, J.J.; Verhagen, H.; Gallagher, A.M. The Potential Application of a Biomarker Approach for the Investigation of Low-Calorie Sweetener Exposure. *Proc. Nutr. Soc.* **2016**, *75*, 216–225. [CrossRef]
58. Ball, L.M.; Renwick, A.G.; Williams, R.T. The Fate of [14C]Saccharin in Man, Rat and Rabbit and of 2-Sulphamoyl[14C]benzoic Acid in the Rat. *Xenobiotica* **1977**, *7*, 189–203. [CrossRef] [PubMed]
59. Collings, A.J. Metabolism of Cyclamate and Its Conversion to Cyclohexylamine. *Diabetes Care* **1989**, *12*, 50–55. [CrossRef]
60. Renwick, A.G.; Thompson, J.P.; O'Shaughnessy, M.; Walter, E.J. The metabolism of cyclamate to cyclohexylamine in humans during long-term administration. *Toxicol. Appl. Pharm.* **2004**, *196*, 367–380. [CrossRef]
61. Grice, H.; Goldsmith, L. Sucralose—An overview of the toxicity data. *Food Chem. Toxicol.* **2000**, *38*, S1–S6. [CrossRef]
62. Wheeler, A.; Boileau, A.; Winkler, P.; Compton, J.; Prakash, I.; Jiang, X.; Mandarino, D. Pharmacokinetics of rebaudioside A and stevioside after single oral doses in healthy men. *Food Chem. Toxicol.* **2008**, *46*, S54–S60. [CrossRef]
63. Huth, P.J.; Fulgoni, V.L.; Keast, D.R.; Park, K.; Auestad, N. Major food sources of calories, added sugars, and saturated fat and their contribution to essential nutrient intakes in the U.S. diet: Data from the national health and nutrition examination survey (2003–2006). *Nutr. J.* **2013**, *12*, 116. [CrossRef]
64. Tasevska, N.; Runswick, S.A.; Welch, A.A.; McTaggart, A.; Bingham, S.A. Urinary sugars biomarker relates better to extrinsic than to intrinsic sugars intake in a metabolic study with volunteers consuming their normal diet. *Eur. J. Clin. Nutr.* **2008**, *63*, 653–659. [CrossRef] [PubMed]
65. Tasevska, N.; Midthune, D.; Potischman, N.; Subar, A.F.; Cross, A.J.; Bingham, S.A.; Schatzkin, A.; Kipnis, V. Use of the Predictive Sugars Biomarker to Evaluate Self-Reported Total Sugars Intake in the Observing Protein and Energy Nutrition (OPEN) Study. *Cancer Epidemiol. Biomark. Prev.* **2011**, *20*, 490–500. [CrossRef]
66. Tasevska, N.; Midthune, D.; Tinker, L.F.; Potischman, N.; Lampe, J.W.; Neuhouser, M.L.; Beasley, J.; Van Horn, L.; Prentice, R.L.; Kipnis, V. Use of a Urinary Sugars Biomarker to Assess Measurement Error in Self-Reported Sugars Intake in the Nutrition and Physical Activity Assessment Study (NPAAS). *Cancer Epidemiol. Biomark. Prev.* **2014**, *23*, 2874–2883. [CrossRef] [PubMed]
67. Tasevska, N. Urinary Sugars—A Biomarker of Total Sugars Intake. *Nutrition* **2015**, *7*, 5816–5833. [CrossRef] [PubMed]
68. Kuhnle, G.G.C.; Tasevska, N.; Lentjes, M.; Griffin, J.L.; Sims, M.A.; Richardson, L.; Aspinall, S.M.; Mulligan, A.A.; Luben, R.; Khaw, K.-T. Association between sucrose intake and risk of overweight and obesity in a prospective sub-cohort of the European Prospective Investigation into Cancer in Norfolk (EPIC-Norfolk). *Public Health Nutr.* **2015**, *18*, 2815–2824. [CrossRef] [PubMed]
69. Bingham, S.; Luben, R.; Welch, A.; Tasevska, N.; Wareham, N.; Khaw, K.T. Epidemiologic Assessment of Sugars Consumption Using Biomarkers: Comparisons of Obese and Nonobese Individuals in the European Prospective Investigation of Cancer Norfolk. *Cancer Epidemiol. Biomark. Prev.* **2007**, *16*, 1651–1654. [CrossRef] [PubMed]
70. Yeung, C.H.; Louie, J.C. Methodology for the assessment of added/free sugar intake in epidemiological studies. *Curr. Opin. Clin. Nutr. Metab. Care* **2019**, *22*, 271–277. [CrossRef]

71. Nash, S.H.; Kristal, A.R.; Bersamin, A.; Choy, K.; Hopkins, S.E.; Stanhope, K.L.; Havel, P.J.; Boyer, B.B.; O'Brien, D.M. Isotopic estimates of sugar intake are related to chronic disease risk factors but not obesity in an Alaska native (Yup'ik) study population. *Eur. J. Clin. Nutr.* **2013**, *68*, 91–96. [CrossRef] [PubMed]
72. O'Brien, D.M. Stable Isotope Ratios as Biomarkers of Diet for Health Research. *Annu. Rev. Nutr.* **2015**, *35*, 565–594. [CrossRef]
73. Rothwell, J.A.; Madrid-Gambin, F.; Garcia-Aloy, M.; Andres-Lacueva, C.; Logue, C.; Gallagher, A.M.; Mack, C.; Kulling, S.E.; Gao, Q.; Praticò, G.; et al. Biomarkers of intake for coffee, tea, and sweetened beverages. *Genes Nutr.* **2018**, *13*, 15. [CrossRef] [PubMed]

3.2 Publication 2

Metabolomics signatures of sweetened beverages and added sugar are related to anthropometric measures of adiposity in young individuals: results from a cohort study. https://doi.org/10.1016/j.ajcnut.2024.07.021

The American Journal of Clinical Nutrition 120 (2024) 879–890

The American Journal of CLINICAL NUTRITION

journal homepage: https://ajcn.nutrition.org/

Original Research Article

Metabolomics signatures of sweetened beverages and added sugar are related to anthropometric measures of adiposity in young individuals: results from a cohort study

Samuel Muli [1,*], Maike E Schnermann [1], Mira Merdas [2], Jodi Rattner [2], David Achaintre [2], Ines Perrar [1], Jantje Goerdten [3], Ute Alexy [1], Augustin Scalbert [2], Matthias Schmid [4], Anna Floegel [3,5], Pekka Keski-Rahkonen [2], Kolade Oluwagbemigun [1,†], Ute Nöthlings [1,†]

[1] *Unit of Nutritional Epidemiology, Department of Nutrition and Food Sciences, University of Bonn, Bonn, Germany;* [2] *International Agency for Research on Cancer (IARC), Lyon, France;* [3] *Department of Epidemiological Methods and Etiological Research, Leibniz Institute for Prevention Research and Epidemiology (BIPS), Bremen, Germany;* [4] *Institute for Medical Biometry, Informatics and Epidemiology (IMBIE), University Hospital Bonn, Bonn, Germany;* [5] *Section of Dietetics, Faculty of Agriculture and Food Sciences, Hochschule Neubrandenburg, Neubrandenburg, Germany*

A B S T R A C T

Background: The associations of sweetened beverages (SBs) and added sugar (AS) intake with adiposity are still debated. Metabolomics could provide insights into the mechanisms linking their intake to adiposity.

Objectives: We aimed to identify metabolomics biomarkers of intake of low- and no-calorie sweetened beverages (LNCSBs), sugar-sweetened beverages (SSBs), and ASs and to investigate their associations with body mass index, body fat percentage, and waist circumference.

Methods: We analyzed 3 data sets from the Dortmund Nutritional and Anthropometric Longitudinally Designed (DONALD) cohort study, of children who provided 2 urine samples ($n = 297$), adolescents who provided a single urine sample ($n = 339$), and young adults who provided a single plasma sample ($n = 195$). Urine and plasma were analyzed using untargeted metabolomics. Dietary intakes were assessed using 3-d weighed dietary records. The random forest, partial least squares, and least absolute shrinkage and selection operator were jointly used for metabolite selection. We examined associations of intakes with metabolites and anthropometric measures using linear and mixed-effects regression.

Results: In adolescents, LNCSB were positively associated with acesulfame (β: 0.0012; 95% confidence interval [CI]: 0.0006, 0.0019) and saccharin (β: 0.0009; 95% CI: 0.0002, 0.0015). In children, the association was observed with saccharin (β: 0.0016; 95% CI: 0.0005, 0.0027). In urine and plasma, SSBs were positively associated with 1-methylxanthine (β: 0.0005; 95% CI: 0.0003, 0.0008; and β: 0.0010, 95% CI 0.0004, 0.0015, respectively) and 5-acetylamino-6-amino-3-methyluracil (β: 0.0005; 95% CI: 0.0002, 0.0008; and β: 0.0009; 95% CI: 0.0003, 0.0014, respectively). AS was associated with urinary sucrose (β: 0.0095; 95% CI: 0.0069, 0.0121) in adolescents. Some of the food-related metabolomics profiles were also associated with adiposity measures.

Conclusions: We identified SBs- and AS-related metabolites, which may be important for understanding the interplay between these intakes and adiposity in young individuals.

Keywords: metabolite biomarkers, metabolomics, sweetened beverages, added sugar, adiposity

Introduction

High consumption of sweetened beverages (SBs) or "soft drinks" and added sugars (ASs), particularly among children and adolescents has emerged as an important nutrition and public health issue [1]. SBs are generally divided into 2 categories, sugar-sweetened beverages (SSBs) and low-calorie and no-calorie sweetened beverages (LNCSBs). SSBs are a major source of ASs in the diet [2] and are

Abbreviations: AAMU, 5-acetylamino-6-amino-3-methyluracil; AS, added sugar; CI, confidence interval; DONALD, Dortmund Nutritional and Anthropometric Longitudinal Designed; IARC, International Agency for Research on Cancer; LASSO, least absolute shrinkage and selection operator; LNCSB, low- and no-calorie sweetened beverage; ML, machine learning; PLS, partial least squares; RF, random forest; SB, sweetened beverage; SDS, standard deviation score; SSB, sugar-sweetened beverage; WC, waist circumference; %BF, body fat percentage; 3d-WDR, 3-d weighed dietary record.

* Corresponding author.
E-mail address: smuli@uni-bonn.de (S. Muli).

† KO and UN contributed equally to this work.

https://doi.org/10.1016/j.ajcnut.2024.07.021
Received 26 January 2024; Received in revised form 11 July 2024; Accepted 22 July 2024; Available online 24 July 2024

argued to contribute to excess caloric intake and poor nutrition [3,4]. Regular consumption of SBs is associated with various health conditions, such as weight gain, obesity, type 2 diabetes, cardiovascular diseases, and some cancers [5–7]. However, some of these associations are not consistent across studies [8,9].

Dietary intake is typically assessed by self-reported dietary questionnaires, which are fraught with measurement errors [10]. Indeed, underreporting of SB and AS intake because of social desirability bias has been described [11]. In recent years, biomarkers of dietary intake have been proposed as one of the ways to improve dietary exposure assessment [12]. To date, however, few reliable biomarkers have been identified and validated for use in epidemiologic studies [13–15]. Some gains include the 24-hour urinary sucrose and fructose, which has been applied in calibrating total sugar intakes in diverse populations [16,17] although it does not distinguish intrinsic from AS. In our previous work [18], we outlined that the ^{13}C:^{12}C carbon isotope ratio ($\delta^{13}C$), measured in whole blood, red blood cells, hair, breath, and plasma correlates with AS and SSB intake. However, sucrose from C4 photosynthetic plants (e.g., corn, sugarcane) moderately correlates with $\delta^{13}C$ compared with sucrose from C3 plants (e.g., sugar beets, most fruits). Thus, the utility of this biomarker is limited by the source of the sucrose [19]. Consistent with the earlier reviews [18,19], an updated review on validity of biomarkers of food intake emphasizes the sustained interest in discovery and validation of new biomarkers, particularly for foods like SSBs [20].

One issue that may have influenced the progress of biomarkers of dietary intake such as for SBs and AS is that many candidates are selected based on putative mechanisms. However, given the aforementioned potentially complex metabolism of these foods, targeting single or multiple selected pathways may be suboptimal. Large-scale metabolite measurement through untargeted metabolomics approaches across multiple data sets and biosamples could help uncover biomarkers of SBs and AS. Further, because metabolites of these foods might exist in a continuum in body fluids, profiling of the plasma and subsequently the urine could be an important research advance. Besides, changes in the metabolome are likely to represent important drivers of the relationship between the intake of intake of SBs and AS and adiposity. Interestingly, limited studies have investigated untargeted metabolomics biomarkers of SSB intake [21–24], as well as the metabolic changes of SSB intake with adiposity [24].

Leveraging 3 data sets across 2 biosamples within a well-characterized cohort of children and adolescents, we aimed to explore metabolomics biomarkers of SBs and AS intake and to investigate their associations with 3 anthropometric measures of adiposity: BMI, body fat percentage (%BF), and waist circumference (WC).

Methods

Study design

The Dortmund Nutritional and Anthropometric Longitudinally Designed (DONALD) study is an open cohort in Dortmund, Germany, that has been recruiting infants in their first year of life since 1985. Participants undergo their first examination at 3 mo of age, followed by 3 additional visits in their first year of life, 2 visits in the second year, and then annually until young adulthood. Regular examinations include dietary intake, anthropometrics, urine samples (starting at age 3–4 y), blood samples (starting at age 18 y), and interviews on lifestyle, sociodemographics, and medical history. A more detailed description of the DONALD study is described elsewhere [25]. The DONALD study was approved by the Ethics Committee of the University of Bonn and conducted according to the guidelines of the Declaration of Helsinki. Written informed consent was obtained from the parents and from adolescents aged 16 y and above.

Study population

This analysis included 3 study samples, hereinafter termed children urine, adolescent urine, and young adult plasma. The eligible participants for children urine were individuals with two 3-d weighed dietary records (3d-WDRs) and 2 urine collections, and for adolescent urine, individuals with one 3d-WDR and 1 urine collection. Young adult plasma comprised individuals with 3 or more 3d-WDR assessments within the last 5 years preceding the date of blood draw. Figure 1 provides an overview of the 3 study samples and the analytical plan. Supplemental Figure 1 provides a detailed flowchart and the overlap of participants across the samples.

Measures

Dietary intake assessment

Study participants themselves, or assisted by their parents, weighed all foods and beverages consumed as well as leftovers using electronic food scales to the nearest 1 g. In situations where weighing is not feasible, for example, out-of-home consumption, participants estimated their intakes semiquantitatively using common household measures (e.g., spoons, cups, portions). Participants provided information on specific food items, their brands, ingredients, and preparation. Trained dieticians entered the dietary records in the database after reviewing them for completeness and plausibility. Using a continuously updated in-house food composition database [26], food group intakes were determined. The nutritional content of staple foods was based on German food composition tables, while the energy and nutrient values of commercial food products were determined by recipe simulation.

Four food groups were included in this analysis: SSB included a diverse group of nondairy beverages with AS, including sweetened fruit juice drinks, sodas, sport energy drinks, and other flavored, carbonated, and noncarbonated soft drinks. LNCSB included aforementioned beverages but artificially sweetened with low- or no-caloric sweeteners, without AS. SBs included both SSB and LNCSB. AS included all sugars added to foods either during processing or manufacturing or during preparation or at the table [27]. The individual average of food intake from 3 recording days in grams per day was calculated.

Anthropometric measurements

Experienced nurses conducted different sets of anthropometric measurements. BMI (in kg/m^2) and WC (centimeters) were determined by standard procedures. In children, BMI standard deviation scores (SDSs) were calculated using the national age- and sex-specific BMI percentiles as per Kromeyer–Hauschild method [28]. %BF was estimated from 4 skinfold thickness measurements (biceps, triceps, iliaca, and scapula), following age- and sex-specific equations by Deurenberg et al. [29]. Anthropometric measurements for this analysis were taken on, or nearest to, the date of urine collection and blood draw.

Other covariates assessment

Habitual leisure time physical activity was assessed using a questionnaire based on the Adolescent Physical Activity Recall Questionnaire [30], considering participation in organized (e.g., club sport, gym) and unorganized sports (e.g., cycling) over the previous year. Energy

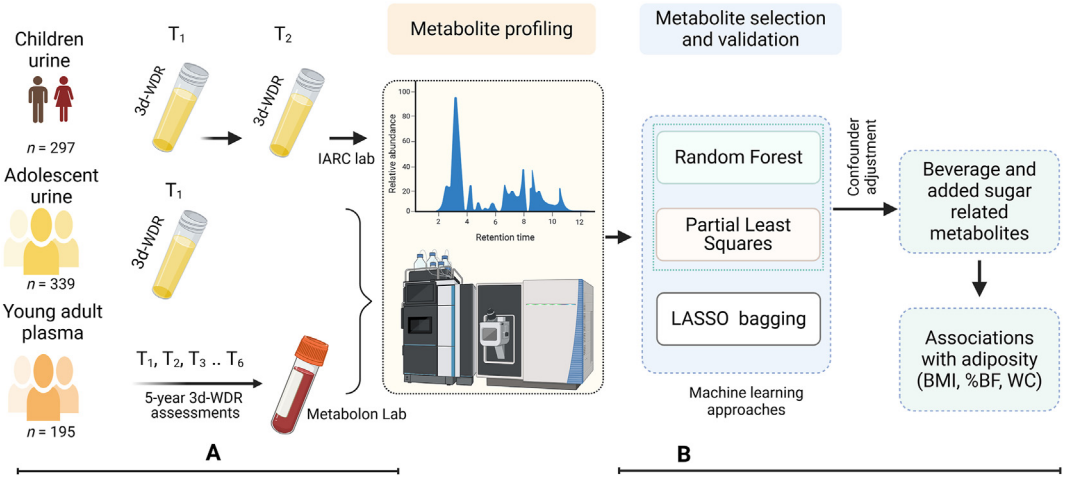

FIGURE 1. Study design and analysis. (A) Analytic samples and data collection. Children urine included 2 dietary assessments and 2 urine collections. Adolescent urine included single dietary assessment and urine collection. Young adult plasma included multiple (3–6) dietary assessments within 5 y preceding a single blood draw. (B) Study analytic plan. %BF, body fat percentage; 3d-WDR, 3-d weighed dietary record; LASSO, least absolute shrinkage and selection operator; T_1–T_6, dietary assessments; WC, waist circumference.

expenditure from these activities was quantified in metabolic equivalent of task hours per week. Self-reported smoking and alcohol status in adolescents and young adults was categorized into current, former, or never. Lifestyle factors (physical activity, smoking, and alcohol use) were assessed in adolescents and young adults only.

Urine samples

The 24-h urine samples were collected on the third day of their dietary assessment, following a standardized protocol. Within this period, urine was collected in preservative-free plastic containers and stored at less than −12°C until transferred to the study center where they were stored at −22°C until thawed and analyzed.

Blood samples

A fasting blood sample was drawn from participants and centrifuged at 4°C for 15 min (3100 U/min, 2000 G). Serum, plasma (citrate, EDTA) and buffy coat aliquots (500 µL each) were subsequently stored at −80°C. EDTA plasma was used in this study.

Metabolite profiling

Two independent laboratories, Metabolon and International Agency for Research on Cancer (IARC), performed untargeted metabolomics analysis, as shown in Figure 1. Metabolon used ultra–high-performance liquid chromatography-tandem mass spectroscopy to identify metabolites in adolescent urine and young adult plasma samples. Briefly, Metabolon carried out a set of standardized procedures from sample accession and preparation to analysis, raw data extraction and peak identification, following their internal standards [31]. In profiling the plasma samples, both metabolomics and lipidomics techniques were applied. In adolescent urine samples, 1407 features were annotated: 940 with known biochemical identity and 467 with no assigned chemical identity. In plasma samples, 1042 features were annotated: 811 with known chemical identity and 231 unknown.

IARC performed untargeted metabolite profiling using a ultra–high-performance liquid chromatography-tandem mass spectroscopy system (Q Exactive; Thermo Fischer Scinetific). Children urine samples (n = 600, representing 2 collections per participant at 2 time points) were analyzed next to each other in random order, and sample pairs were randomized across the batch. There were 4 independent analytical batches consisting of 2 individual 96-well plates. The mass spectrometer was operated in a positive/negative switching polarity. Preprocessing was performed using Compound Discoverer 3.3 software (Thermo Fisher Scientific) with minimum peak intensity threshold at 500,000, mass tolerance at 5 ppm, and feature alignment between samples performed with maximum retention time window of 0.05 min and mass tolerance of 5 ppm. Unlike Metabolon's approach, metabolite features from IARC were first analyzed with dietary intake, and only features related to dietary intake were subsequently annotated. A detailed description of the analytical, quality control standards, and annotation procedures for both laboratories, is provided in Supplemental Methods.

Statistical analyses

Participant characteristics

We calculated the median (25% and 75% percentile) for continuous variables and count (percentage) for categorical variables.

Preprocessing metabolomics data

We excluded metabolites that had missing values in >30% of the consumers of each food group. Missing values were imputed with half of the minimum value observed within each batch, with the assumption of missing due to low concentration below the instrument's detection limit. Subsequently, these were natural log-transformed and standardized to have a mean of zero and unit variance. We corrected analytical batch effects by *ber* bagging method using the 'dbnorm' R package [32].

Metabolites selection

We applied 3 machine learning (ML) methods to first select and validate food-related metabolites, acknowledging the high dimensionality of the data sets and correlation among metabolites. These were random forest (RF), partial least squares (PLS), and least absolute shrinkage and selection operator (LASSO) with a bagging strategy. The PLS and RF were implemented using the multivariate modeling with minimally biased variable selection in R algorithm, a statistical validation framework that integrates a recursive ranking and backward

elimination of variables within a repeated double cross-validation scheme [33]. The models were tuned following author recommendations [33] and were repeated 50 times to identify a stable set of metabolites ranked based on their importance to predict the respective dietary variable. The LASSO models were implemented using a variable-selection oriented LASSO bagging algorithm, combining LASSO regression with bootstrap aggregating to enhance stability and robust selection of biomarkers [34]. We generated 1000 bootstrap samples from the original data sets, and LASSO models were fitted on each sample using 5-fold cross-validation, all other parameters as per author description [34]. We applied the curve elbow point method to detect sharp drops in the observed frequency of variable selection. Metabolites with selection frequencies at and above the last elbow point were retained, if more than 1 point existed. For downstream analyses, we considered only metabolites selected by ≥2 ML methods to reduce the likelihood of selecting metabolites due to noise or method-specific bias. An overview of these steps is provided in Supplemental Figure 2.

Associations of food groups with metabolites

We used multivariable linear regression and linear mixed-effects models to examine the association of dietary intake and individual metabolites, for cross-sectional and repeated measures, respectively. In all regression models, we regressed each metabolite on intake (grams per day) adjusting for age, sex, and energy intake. For adolescents and young adults, we further adjusted for lifestyle factors (physical activity, alcohol, and smoking status). In children urine samples, the linear mixed-effects models included a random intercept for each participant. Because of the analytic design of long-term dietary assessment, plasma models were additionally adjusted for the difference in time between dietary assessment and blood draw (i.e., difference = age at blood draw – mean age of dietary assessments) and the number of dietary assessments. To account for multiple testing, we applied the Benjamini–Hochberg procedure to control the false discovery rate at 5%.

Associations of food-related metabolites with anthropometric measurements

To assess the associations of the food-related metabolites and adiposity, separate linear regression and linear mixed-effects models were constructed for each adiposity measure (BMI, %BF, and WC). We modeled these as response variables and sets of food-related metabolites and covariates [age, sex, energy intake, birthweight, and time difference (in days) between biosample collection and anthropometric measurements, and additionally, in adolescents and young adults, physical activity, alcohol, and smoking status] as predictor variables. In children urine samples, a random intercept for each participant was specified. To assess multicollinearity of the predictor variables, we used the variance inflation factor, and whenever appropriate, removed redundant metabolites with variance inflation factor of >10 [35], progressively starting from the highest. Considering the strong correlation between anthropometric measurements, we applied the modified Bonferroni method [36] to adjust the significance level for multiple testing.

Missing covariates

We used the K-Nearest Neighbor algorithm to impute the missing values in birthweight, physical activity, alcohol, and smoking status, with 10 nearest neighbors based on nonmissing values in other covariates (sex, age, BMI, energy intake, birthweight, physical activity, and alcohol and smoking status) implemented in the VIM R package. All statistical analyses were conducted using R 4.1.3 (The R Foundation for Statistical Computing).

Results

Participant characteristics

The median ages at biosample collection were 7.0 y (T_1) and 8.0 y (T_2) for children, 18.0 y for adolescents, and 18.1 y for young adults. Sex distribution was 52.9% female for T_1 and 51.5% for T_2 in children; 49.0% for adolescents; and 55.4% for young adults (Table 1).

Metabolite selections

There was good agreement in metabolite selections across the ML approaches. The PLS consistently selected more metabolites and shared common selections with the LASSO bagging algorithm, compared with the RF. The metabolite selections are provided for children urine, adolescent urine, and young adult plasma in Supplemental Tables 1–3, respectively.

Associations of food groups with urine metabolites in children

LNCSB, SSB, SBs, and AS were associated with 4, 18, 18, and 28 metabolite features, respectively (Supplemental Table 4). Of the 8 biochemically identified metabolites, 7 associations were food specific (LNCSB positively associated with saccharin; SSB negatively associated with 4-pyridoxic acid; SBs negatively associated with kynurenic acid; and AS positively associated with theobromine, 7-methylguanine, aspartyphenylalanine, and negatively associated with uric acid). There was 1 common food-metabolite association: SSB, SBs, and AS were all associated with higher urinary concentrations of decadienoyl carnitine C10:2 (Table 2).

Associations of food groups with urine metabolites in adolescents

Figure 2 summarizes metabolites associated with SBs and AS. LNCSB intake was associated with 11 metabolites (6 of them were specific to LNCSB), SSB intake with 37 metabolites (12 specific), SBs intake with 34 metabolites (9 specific), and AS intake with 32 metabolites (24 specific). All food-specific and nonspecific associations in adolescent urine samples can be found in Supplemental Table 5.

Briefly, LNCSB intake was associated with higher concentrations of acesulfame (β: 0.0012; 95% confidence interval [CI]: 0.0006, 0.0019) and saccharin (β: 0.0009; 95% CI: 0.0002, 0.0015). SSB intake was associated with higher concentrations of caffeine metabolites: 1-methylxanthine (β: 0.0005; 95% CI: 0.0003, 0.0008) and 5-acetylamino-6-amino-3-methyluracil (AAMU; β: 0.0005; 95% CI: 0.0002, 0.0008). Notably, SSB and SB intakes were also associated with elevated concentrations of unknown metabolites X-17679 (β: 0.0010; 95% CI: 0.0008, 0.0013; and β: 0.0010; 95% CI: 0.0007, 0.0012); X-19497 (β: 0.0005; 95% CI: 0.0003, 0.0008; and β: 0.0005; 95% CI: 0.0003, 0.0008); and X-17328 (β: 0.0005; 95% CI: 0.0003, 0.0008; and β: 0.0006; 95% CI: 0.0003, 0.0008), respectively. Other noteworthy associations included N1-methyl-2-pyridone-5-carboxamide (2PYr) with SSBs (β: 0.0004; 95% CI: 0.0002, 0.0007) and SBs (β: 0.0004; 95% CI: 0.0002, 0.0007) and N,N-dimethylalanine with SSBs (β: 0.0005; 95% CI: 0.0002, 0.0007) and SBs (β: 0.0005; 95% CI: 0.0003, 0.0008). AS intake was associated with higher concentrations of sucrose (β: 0.0095; 95% CI: 0.0069, 0.0121), X-17679 (β: 0.0098; 95% CI:

TABLE 1
Basic characteristics of the study population[1]

	Children urine				Adolescent urine		Young adult plasma	
	n	T_1, *n* = 297	*n*	T_2, *n* = 270	*n*	Urine, *n* = 339	*n*	Plasma, *n* = 195
Sex: female	297	157 (52.9)	270	139 (51.5)	339	166 (49.0)	195	108 (55.4)
Age at biosample collection (y)	297	7.0 (7.0, 7.2)	270	8.0 (8.0, 8.2)	339	18.0 (17.0, 18.1)	195	18.1 (18.1, 18.2)
BMI (kg/m^2)	297	15.8 (15.0, 17.1)	270	16.2 (15.1, 17.5)	339	21.9 (19.9, 24.0)	195	22.2 (20.1, 24.5)
Body fat percentage	296	17.3 (14.7, 20.4)	270	17.5 (14.8, 21.1)	339	22.6 (18.4, 27.1)	195	23.6 (19.2, 28.1)
LNCSB (g/d)	297	0.0 (0.0, 1022.3)	270	0.0 (0.0, 443.3)	339	0.0 (0.0, 0.0.0)	195	0.0 (0.0, 54.2)
Sugar-sweetened beverage (g/d)	297	44.0 (0.0, 163.3)	270	55.8 (0.0, 166.9)	339	133.3 (0.0, 418.3)	195	124.9 (51.2, 324.4)
Total sweetened beverages (g/d)	297	66.7 (0.0, 198.3)	270	66.7 (0.0, 216.9)	339	166.7 (0.0, 508.3)	195	163.8 (72.8, 408.4)
Added sugar (g/d)	297	46.5 (33.7, 64.7)	270	49.9 (35.0, 72.3)	339	62.2 (35.7, 89.9)	195	62.2 (43.8, 82.2)
Added sugar (% energy)	297	12.5 (9.4, 16.3)	270	12.5 (9.1, 17.2)	339	11.6 (7.4, 16.0)	195	12.5 (9.6, 15.3)
TEI (kcal/d)	297	1527.3 (1310.0, 1736.2)	270	1635.2 (1402.2, 1840.3)	339	2126.9 (1748.5, 2582.1)	195	1978.1 (1697.0, 2390.1)
Dietary assessments	297	1.0	270	1.0	339	1.0	195	4.0 (4.0, 5.0)
Physical activity (MET-h/w)	—	—	—	—	215	34.0 (14.1, 54.8)	184	30.1 (12.1, 52.9)
Smoking status	—	—	—	—	211		142	
Never	—	—	—	—	—	155 (73.5)	—	98 (69.0)
Former	—	—	—	—	—	23 (10.9)	—	21 (14.8)
Current	—	—	—	—	—	33 (15.6)	—	23 (16.2)
Alcohol status	—	—	—	—	179		153	
Never	—	—	—	—	—	24 (13.4)	—	20 (13.1)
Former	—	—	—	—	—	27 (15.1)	—	31 (20.3)
Current	—	—	—	—	—	128 (71.5)	—	102 (66.7)

Abbreviations: LNCSB, low- and no-calorie sweetened beverages; MET-h/w, metabolic equivalent of task-hours per week; TEI, total energy intake.
[1] Data are given as *n* (%) and median (25%, 75%) for categorical and continuous variables, respectively. In children analytic sample, of the 297 participants in T_1, 270 had repeated measures (T_2). Although blood samples are collected at the age of 18 y or older, the dietary assessments in "young adults" mostly occurred during adolescence. Differences in *n* are due to missing data.

TABLE 2
Multivariable linear regression estimates of the associations of food groups with urine metabolites in children (*n* = 297)

Food	Metabolite	HMBD ID	β	95% CI	
				Lower	Upper
LNCSB	Saccharin	HMDB0029723	0.0016	0.0005	0.0027
SSB	Decadienoyl carnitine (C10:2)		0.0014	0.0009	0.0019
	4-pyridoxic acid	HMDB0000017	−0.0006	−0.0011	−0.0002
SBs	Decadienoyl carnitine (C10:2)		0.0013	0.0008	0.0017
	Kynurenic acid	HMDB0000715	−0.0008	−0.0013	−0.0003
AS	Decadienoyl carnitine (C10:2)		0.0120	0.0085	0.0155
	Theombromine	HMDB0002825	0.0080	0.0044	0.0116
	7-Methylguanine	HMDB000089	0.0055	0.0020	0.0092
	Aspartylphenylalanine	HMDB0000706	0.0050	0.0016	0.0084
	Uric acid	HMDB0000289	−0.0038	−0.0072	−0.0003

Abbreviations: AS, added sugar; CI, confidence intervals; HMBD ID, human metabolome database identification; LNCSB, low- and no-calorie sweetened beverages; SB, total sweetened beverage; SSB, sugar-sweetened beverage.
Models adjusted for age, sex, and energy intake, with a random intercept for each participant. Only biochemically identified metabolites with false discovery rate–adjusted *q* value <0.05 are shown. Complete list is given in Supplemental Table 1.

0.0073, 0.0124), and 1-methylxanthine (β: 0.0084; 95% CI: 0.0057, 0.0111).

Associations of food groups with plasma metabolites

The associations of SBs and AS intakes with plasma metabolites are provided in Table 3. LNCSB intake was associated with 3 metabolites (2 of them specific to LNCSB intake), SSB with 11 metabolites (5 specific), SBs with 15 metabolites (8 specific), and AS with 3 metabolites (1 specific). Notably, 1-methyxanthine and AAMU were positively associated with SSBs (β: 0.0010; 95% CI: 0.0004, 0.0015; and β: 0.0009; 95% CI: 0.0003, 0.0014); SBs (β: 0.0010; 95% CI: 0.0005, 0.0015; and β: 0.0008; 95% CI: 0.0003, 0.0013); and AS (β: 0.0089; 95% CI: 0.0029, 0.0150; and β: 0.0091; 95% CI: 0.0031, 0.0151), respectively. Moreover, SSBs and SBs were associated with higher concentrations of caffeine (β: 0.0010; 95% CI: 0.0004, 0.0015; and β: 0.0009; 95% CI: 0.0005, 0.0014) and 1-3-dimethylurate (β: 0.0009; 95% CI: 0.0003, 0.0014; and β: 0.0008; 95% CI: 0.0004, 0.0013), respectively. All food-specific and nonspecific associations in plasma can be found in Supplemental Table 6.

Associations of food-related metabolites with anthropometric measures

In children, 4 AS-related metabolite features, but of unknown biochemical identities, had mixed associations with BMI SDS and %

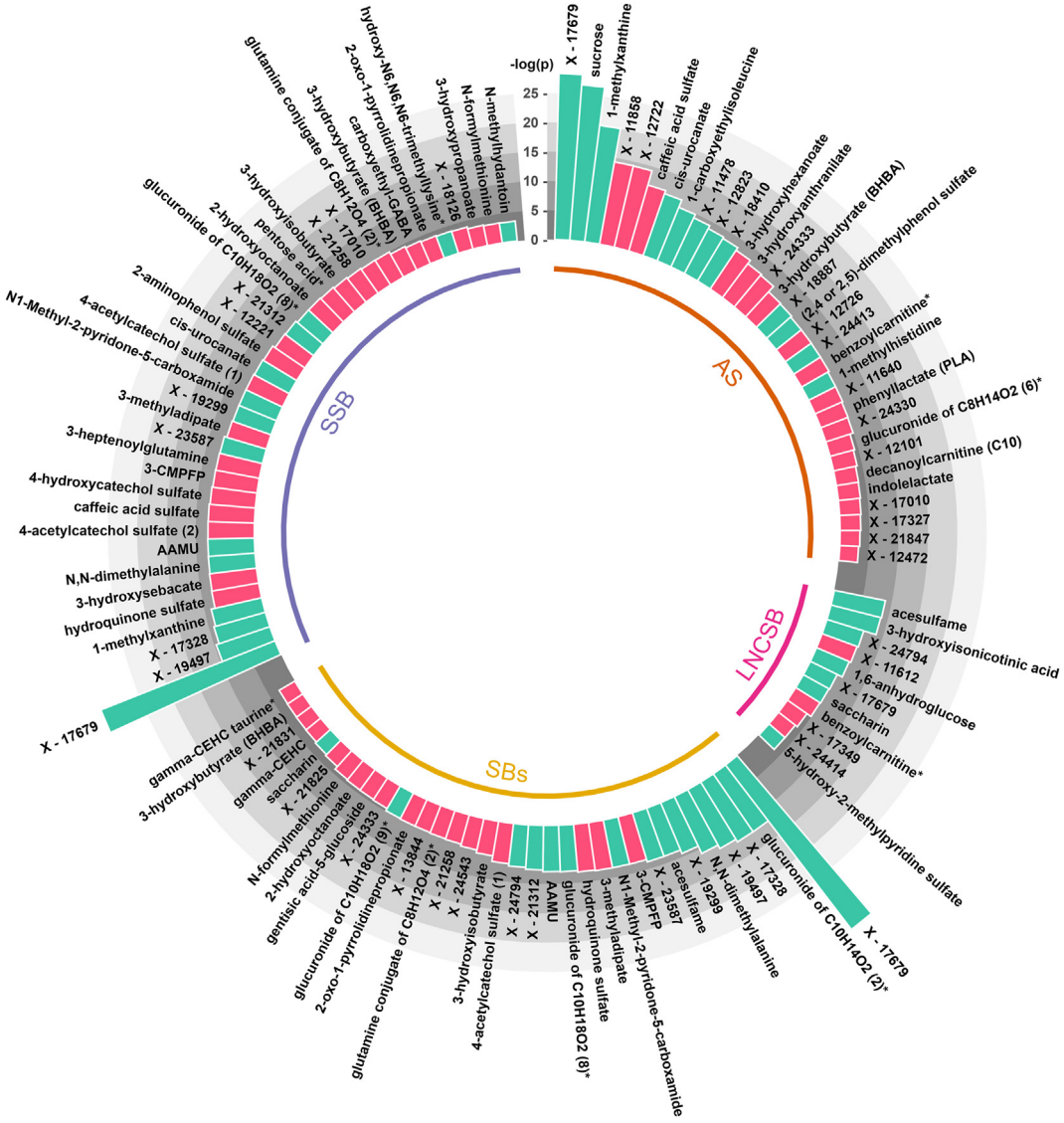

FIGURE 2. The associations of food groups with urine metabolites in adolescents. All models were adjusted for age, sex, energy intake, physical activity, alcohol and smoking status. Metabolites with false discovery rate–adjusted q value <0.05: AS, $n = 32$; LNCSB, $n = 11$; SSB, $n = 37$, and SBs, $n = 34$. The histogram bars represent the log-transformed P values: turquoise, positive association; light red, negative. *Metabolites not confirmed based on authentic standard, but Metabolon are confident in its identity. AS, added sugar; LNCSB, low- and no-calorie sweetened beverage; SB, total sweetened beverage; SSB, sugar-sweetened beverage.

BF (Table 4). One metabolite feature "214.08427@3.876" was positively associated with both BMI SDS (β: 0.08; 95% CI: 0.01, 0.014) and %BF (β: 0.59; 95% CI: 0.16, 1.02).

In adolescent urine samples, acesulfame was positively associated with BMI (β: 0.82; 95% CI: 0.43, 1.22) and %BF (β: 0.92; 95% CI: 0.36, 1.47). 2PYr was positively associated with BMI (β: 0.63; 95% CI: 0.21, 1.05), %BF (β: 0.86; 95% CI: 0.27, 1.44), and WC (β: 1.98; 95% CI: 0.64, 3.32). Decanoylcarnitine (C10) was also positively associated with BMI (β: 1.09; 95% CI: 0.62, 1.56), %BF (β: 1.56; 95% CI: 0.92, 2.20), and WC (β: 4.41; 95% CI: 2.85, 5.96). Two metabolites were inversely associated with all adiposity measures: N,N-dimethylalanine with BMI (β: −0.61; 95% CI: −1.03, −0.18); %BF (β: −1.18; 95% CI: −1.78, −0.58); and WC (β: −2.27; 95% CI: −3.66, −0.89); and 3-hydroxyhexanoate with BMI (β: −0.96; 95% CI: −1.49, −0.42); %BF (β: −1.43; 95% CI: −2.15, −0.71); and WC (β: −3.14; 95% CI: −4.86, −1.42) (Table 4).

In young adult plasma samples, 2 unknown metabolites showed positive associations: X-17340 with BMI (β: 1.02; 95% CI: 0.31, 1.74) and %BF (β: 1.26; 95% CI: 0.26, 2.26); and X-24337 with BMI (β: 0.76; 95% CI: 0.18, 1.33) only. Carotene diol (1) was inversely associated with WC (β: −2.04; 95% CI: −3.50, −0.58) (Table 4).

Post-hoc exploration and analysis of bias

We sought to understand the overlapping of caffeine-related metabolites with SSB and AS across biosamples and whether these associations were confounded by other dietary sources of caffeine, such as coffee, chocolate, and other powdered instant beverages, including tea. Indeed, caffeine and caffeine-related metabolites measured in our study were associated with coffee intake, further providing plausibility of our results (Supplemental Table 7), but the main findings were robust to further adjustment with these foods (Supplemental Table 8). Next, we determined the correlation between SSB and AS intake and found a moderately strong Pearson correlation in adolescent urine (r: 0.65; 95% CI: 0.58, 0.71) and in young adult plasma (r: 0.78; 95% CI: 0.72, 0.83) samples.

TABLE 3
Multivariable linear regression estimates of the associations of food groups with plasma metabolites ($n = 195$)

Food	Metabolite	HMBD ID	β	95% CI	
				Lower	Upper
LNCSB	Octadecanedioylcarnitine (C18-DC)[1]		−0.0022	−0.0034	−0.0009
	Adipoylcarnitine (C6-DC)	HMDB61677	−0.0022	−0.0035	−0.0008
	3-bromo-5-chloro-2,6-dihydroxybenzoic acid[1]		0.0020	0.0007	0.0033
SSB	1-methylxanthine	HMDB10738	0.0010	0.0004	0.0015
	Caffeine	HMDB01847	0.0010	0.0004	0.0015
	1,3-dimethylurate	HMDB01857	0.0009	0.0003	0.0014
	AAMU	HMDB04400	0.0009	0.0003	0.0014
	X-16087		−0.0008	−0.0014	−0.0003
	3-CMPFP	HMDB61643	−0.0008	−0.0014	−0.0002
	X-13866		−0.0007	−0.0013	−0.0002
	Cyclopropyl 10:1 fatty acid (1)[1]		−0.0007	−0.0013	−0.0001
	Carotene diol (1)		−0.0006	−0.0011	−0.0001
	4-cholesten-3-one	HMDB00921	−0.0007	−0.0013	−0.0001
	X-24669		0.0007	0.0001	0.0012
SBs	1-methylxanthine	HMDB10738	0.0010	0.0005	0.0015
	Caffeine	HMDB01847	0.0009	0.0005	0.0014
	AAMU	HMDB04400	0.0008	0.0003	0.0013
	1,3-dimethylurate	HMDB01857	0.0008	0.0004	0.0013
	X-24951		0.0008	0.0003	0.0012
	X-16087		−0.0007	−0.0012	−0.0003
	X-24337		0.0007	0.0003	0.0012
	3-CMPFP	HMDB61643	−0.0008	−0.0012	−0.0003
	X-11308		0.0007	0.0002	0.0012
	Hydroquinone sulfate	HMDB02434	−0.0007	−0.0012	−0.0002
	X-17340		−0.0006	−0.0010	−0.0001
	N-formylphenylalanine		−0.0006	−0.0011	−0.0001
	Dihomo-linolenoylcarnitine (C20:3n3 or 6)[1]		0.0005	0.0001	0.0009
	Adipoylcarnitine (C6-DC)	HMDB61677	−0.0006	−0.0011	−0.0001
	Glutamine conjugate of C6H10O2 (1)[1]		−0.0006	−0.0010	−0.0001
AS	Etiocholanolone glucuronide	HMDB04484	−0.0115	−0.0177	−0.0053
	AAMU	HMDB04400	0.0091	0.0031	0.0151
	1-methylxanthine	HMDB10738	0.0089	0.0029	0.0150

Abbreviations: 3-CMPFP, 3-carboxy-4-methyl-5-pentyl-2-furanpropionate; AAMU, 5-acetylamino-6-amino-3-methyluracil; AS, added sugar; HMBD ID, human metabolome database identification; LNCSB, low- and no-calorie sweetened beverage; SB, total sweetened beverage; SSB, sugar-sweetened beverage. All models adjusted for age, sex, energy intake, physical activity, alcohol and smoking status, number of dietary assessments, and the difference in time between dietary assessment and blood draw. Only metabolites with false discovery rate–adjusted q value < 0.05 are shown. The identities of X-, followed by a number (e.g., X-16087), are unknown.

[1] Metabolites not confirmed based on an authentic standard, but Metabolon are confident in its identity.

Further, we examined potential bias due to half-minimum imputation of metabolite data in our analysis, as suggested previously [37–39]. We compared our results with those from quantile regression imputation, argued to be most optimal for limit of detection missingness [37] and RF imputation, favored for missing completely at random [39]. We observed comparable results across these imputation methods (Supplemental Figure 3).

Finally, we investigated the robustness of the observed associations of food-related metabolites with adiposity in adolescent urine and plasma using a different approach. For each anthropometric measurement, confounder-adjusted food-related metabolites were jointly fit in adaptive elastic-net regularized linear regression models as described previously [40]. We demonstrate that our main findings were robust and invariant to statistical modeling approach (Supplemental Tables 9 and 10).

Discussion

This epidemiologic investigation, using 3 analytic data sets and both urine and plasma samples, identified robust metabolomics biomarkers of SBs and AS. In this study, we confirmed some previously reported metabolite biomarkers of SBs and AS and, to our knowledge, uncovered new ones that are robust across analytic samples. We also observed food-related metabolites that were consistently related to multiple anthropometric measures of adiposity.

Our children and adolescent data showed that urinary acesulfame and saccharin reflect LNCSB intake, with saccharin robust in both analytic samples. Given that LNCSB represent one of the main dietary sources of artificial sweeteners [41], acesulfame and saccharin are plausible urinary metabolite biomarkers of LNCSB intake. The 2 metabolites share similar biochemical properties, including absorption, distribution, metabolism, and excretion [42,43]. We did not detect the other artificial sweeteners in our urine samples. Their specific metabolism and excretion pathways may explain this result: metabolism into other compounds diluted in a large plasma/urine pool (e.g., aspartame into aspartic acid and phenylalanine), not detected by our analytical methods (e.g., steviol glycosides into glucuronides), or not absorbed in the gut (e.g., sucralose) [42,43]. Taken together, these results also suggest that our findings for other food-related metabolites are unlikely to be spurious and their relationship is metabolically plausible.

Moreover, our results indicated that caffeine metabolites, particularly 1-methylxanthine and AAMU, are consistently associated with SSB intake in adolescent urine and plasma samples, independent of all

TABLE 4
Associations of food-related metabolites with adiposity measures

Food-metabolite	β	95% CI		*P*	β	95% CI		*P*	β	95% CI		*P*
		Lower	Upper			Lower	Upper			Lower	Upper	
Children urine	BMI SDS, $n = 297$				%BF, $n = 297$				WC			
165.07939@2.148	0.04	0.01	0.08	0.0092[1]	0.06	−0.17	0.30	0.5927	—	—	—	—
214.08427@3.876	0.08	0.01	0.14	0.0156[1]	0.59	0.16	1.02	0.0078[1]	—	—	—	—
153.04277@2.289	−0.03	−0.06	0.00	0.0515	−0.28	−0.51	−0.05	0.0176[1]	—	—	—	—
166.04911@1.902	−0.10	−0.18	−0.01	0.0293	−0.77	−1.39	−0.15	0.0158[1]	—	—	—	—
Adolescent urine	BMI, $n = 339$				%BF, $n = 339$				WC, $n = 231$			
X-24333	1.31	0.70	1.92	<0.0001[1]	1.10	0.25	1.95	0.0113[1]	2.31	0.20	4.42	0.0321
Acesulfame	0.82	0.43	1.22	0.0001[1]	0.92	0.36	1.47	0.0013[1]	1.40	0.10	2.70	0.0352
2PYr	0.63	0.21	1.05	0.0034[1]	0.86	0.27	1.44	0.0042[1]	1.98	0.64	3.32	0.0040[1]
N,N-dimethylalanine	−0.61	−1.03	−0.18	0.0055[1]	−1.18	−1.78	−0.58	0.0001[1]	−2.27	−3.66	−0.89	0.0014[1]
X-17679	−0.67	−1.20	−0.14	0.0128[1]	−0.41	−1.15	0.33	0.2721	−2.47	−4.15	−0.80	0.0040[1]
X-17010	−0.56	−1.00	−0.11	0.0139[1]	−0.63	−1.25	−0.01	0.0470	−1.13	−2.56	0.30	0.1198
X-17328	0.51	0.09	0.93	0.0176[1]	0.59	−0.00	1.18	0.0504	1.13	−0.16	2.43	0.0856
Decanoylcarnitine (C10)	1.09	0.62	1.56	<0.0001[1]	1.56	0.92	2.20	<0.0001[1]	4.41	2.85	5.96	<0.0001[1]
3-hydroxyhexanoate	−0.96	−1.49	−0.42	0.0005[1]	−1.43	−2.15	−0.71	0.0001[1]	−3.14	−4.86	−1.42	0.0004[1]
γ-CEHC taurine[2]	−1.15	−1.89	−0.40	0.0026[1]	−1.57	−2.58	−0.56	0.0024[1]	−2.65	−4.90	−0.41	0.0209
X-18887	−0.77	−1.36	−0.18	0.0103[1]	−0.67	−1.47	0.13	0.1018	−2.15	−3.92	−0.37	0.0179[1]
Glucuronide of C8H14O2 (6)[2]	0.57	0.02	1.12	0.0422	1.17	0.43	1.91	0.0022[1]	2.44	0.68	4.20	0.0069[1]
X-24330	0.52	0.02	1.02	0.0434	0.85	0.16	1.53	0.0152[1]	0.77	−0.93	2.46	0.3731
X-13844	−0.30	−0.78	0.18	0.2210	−0.79	−1.44	−0.13	0.0188[1]	−1.33	−2.79	0.14	0.0751
Cis-urocanate	−0.41	−0.82	−0.00	0.0488	−0.21	−0.78	0.36	0.4678	−1.55	−2.82	−0.29	0.0164[1]
Young adult plasma	BMI, $n = 195$				%BF, $n = 195$				WC, $n = 195$			
X-17340	1.02	0.31	1.74	0.0053[1]	1.26	0.26	2.26	0.0135[1]	2.05	0.31	3.79	0.0209
X-24337	0.76	0.18	1.33	0.0101[1]	0.42	−0.38	1.22	0.2991	1.52	0.13	2.92	0.0326
Carotene diol (1)	−0.67	−1.27	−0.07	0.0279	−0.73	−1.57	0.10	0.0855	−2.04	−3.50	−0.58	0.0063[1]

Abbreviations: %BF, body fat percentage; 2PYr, N1-methyl-2-pyridone-5-carboxamide; SDS, standard deviation score; WC, waist circumference.
Children urine: Adjusted for age, sex, energy intake, birthweight, and time difference between biosample collection and anthropometric measurements (in their original scale). WC measurements not available. Adolescent urine included all confounder adjustments for children samples, plus physical activity, smoking, and alcohol status. Young adult plasma included all confounder adjustments for adolescent urine samples, plus time difference between dietary assessment and blood draw and number of dietary assessments. The identities of X, followed by a number (e.g., X-24333), and the format "165.07939@2.148" are unknown.
[1] Significant results: children urine, $P < 0.0211$; adolescent urine, $P < 0.0205$; and young adult plasma $P < 0.0199$ (modified Bonferroni method). Only food-related metabolites considered significant with either of the adiposity measures are shown.
[2] Metabolites not confirmed based on authentic standard, but Metabolon are confident in its identity.

other plausible sources of caffeine as shown in our sensitivity analysis. Our study, therefore, confirms the association of SSB intake and elevated concentrations of AAMU in plasma [44] and additionally reports its reflection in urine.

A previous study proposed that SSB ingredients or their combinations could be explored as potential biomarkers for SSB as a group or subtypes of SSBs [19]. Although caffeine is one of the main ingredients of most SSBs, and multiple caffeine metabolites were consistently uncovered in adolescent urine and plasma, an important question remains as to whether a single metabolite biomarker of caffeinated SSB is possible. Based on our findings, it appears that a more promising approach to advance this science should consider combining metabolite biomarkers. We suggest that AAMU and 1-methylxanthine are promising urine and plasma metabolite biomarkers for caffeinated SSB and could be considered alongside other biomarkers such as isotopic signature $\delta^{13}C$ or metabolite biomarkers of SSB ingredients such as taurine [21]. Besides, the SSBs are diverse, with varied concentrations of caffeine, taurine, and other ingredients, making it unlikely that a single, ingredient-based metabolite could reliably reflect overall SSB intake. The unknown biochemical compounds X-17679, X-19497, and X-17328, associated with SSB intake, represent an additional challenge.

Our untargeted metabolomics approach also confirmed the well-established association between AS intake and 24-hour urinary sucrose, reported in targeted approaches [16,45,46]. Our AS variable reflects intakes from various dietary sources. Despite a substantial portion originating from SSB and sugary snacks (e.g., cakes, candies, and desserts), 24-h urinary sucrose does not discriminate specific sources and would not be an ideal biomarker for SSB. This limitation of urinary sucrose, as well as of the isotopic signature $\delta^{13}C$, is extensively discussed elsewhere [47].

Besides the aforementioned putative metabolite biomarkers of SBs and AS, we also uncovered other metabolomics profiles worth highlighting. In children, SSB, SBs, and AS intake correlated with higher concentrations of decadienoyl carnitine (C10:2), a medium-chain acyl-carnitine involved in energy metabolism pathways [48]. We note that medium-chain acyl-carnitines are increasingly investigated as links to various metabolic dysfunctions [48–50] and depression [51]. To our knowledge, the association of C10:2 with SBs and AS intake has not been reported, but elevated concentrations of C10:2 with pork intake have been described [52].

The association between AS and aspartylphenylalanine may reflect the biochemical conversion of aspartame into aspartyl, phenylalanine, and methanol [53], and could indirectly relate to the positive correlation between SBs (sweetened with aspartame) and AS intakes as shown in our sensitivity analysis. The underlying mechanism of AS intake and elevated urinary 7-methylguanine, a biomarker of DNA damage and metabolic rate [54] is unclear. However, in another study, higher concentrations of 7-methylguanine were associated with unhealthy dietary habits [54].

Lower concentrations of kynurenic acid with intake of certain foods, such as SBs in our study, has been described in a longitudinal study [55], and this association has been observed in western-style dietary pattern [56]. Of note, kynurenic acid is an important metabolite of the tryptophan–kynurenine pathway, which is involved in modulation of inflammation and oxidative stress [57].

In adolescent urine and young adult plasma samples, SBs were associated with lower concentrations of 3-carboxy-4-methyl-5-pentyl-2-furanpropanoic acid, a metabolite of furan fatty acids. Humans acquire dietary furan fatty acids mainly from fish and fish oil [58] and are metabolized into 2 major metabolites: 3-carboxy-4-methyl-5-pentyl-2-furanpropanoic acid and 3-carboxy-4-methyl-5-propyl-2-furanpropanoic acid. Our findings across biosamples are therefore of interest, considering the important role furan fatty acids and health [58–60]. It is unclear whether there exists any interaction between SBs and furan fatty acid metabolism or this association is simply because dietary patterns characterized by higher intakes of SBs and AS correlate with overall poor diet quality [61,62]. In parallel, in adolescent urine and young adult plasma samples, we also observed an inverse association between SBs and hydroquinone sulfate, a specific marker of pear intake [63]. Pears are rich in dietary fibers, antioxidative flavonoids, and anti-inflammatory properties [64].

Regarding food-metabolite associations with adiposity, 4 metabolites in children urine samples showed mixed direction of associations with BMI and %BF. Their biochemical identities could not be identified. In adolescent urine samples, acesulfame was positively associated with both BMI and %BF. Acesulfame is not only a common sweetener for beverages but also added in confectionery, sweet, and savory snacks. Our sensitivity analysis showed poor correlation between acesulfame and self-reported intakes of these food items. Thus, we considered that LNCSBs were the likely primary source. This association could also suggest reverse causation and residual confounding, wherein individuals consuming LNCSB may already be overweight, and their beverage choices may be motivated by the intention to lose weight or to restrict their energy intake [65].

Similarly, the 2PYr concentrations were elevated with higher SSB and SBs intakes and positively with BMI, %BF, and WC. SBs are fortified with niacin, whose main metabolites are 2PYr and N-1-methylnicotinamide. Beneficial effects of niacin include neuroprotection, anti-inflammation, and immune modulation [66]. However, short-term metabolic effects of overconsumption of fortified beverages, such as glucose metabolism insulin secretion, have been observed in adolescents [67]. Their long-term effects on adiposity warrant further investigation.

Decanoylcarnitine (C10), positively associated with all adiposity measures in our study, is one of the medium-chain acyl-carnitines linked to body weight [68]. N,N-dimethylalanine and 3-hydroxyhexanoate were inversely associated with all adiposity measures; however, their biological basis remains unclear.

In plasma, carotene diol, a marker of leafy green and cruciferous vegetable intake [20], showed an inverse association with WC. This is consistent with findings from a large cohort study, where serum carotenoids correlated negatively with visceral adiposity [69]. We note that carotenoids are involved in oxidative and lipid metabolism [69] and higher concentrations of carotenoids are favorable for metabolic health. The biological role of X-17340 (associated with higher BMI and %BF) and X-24337 (higher BMI) are unknown.

This study also contributes to the public health discourse on caffeine and sugar pairing and health risks [70,71], by showing that caffeine added to SBs is also reflected at molecular level. Two randomized controlled trials showed that co-ingestion of carbohydrate load and caffeine impaired glucose and insulin responses in young, healthy males [72] and caffeine-containing energy drinks and shots resulted in acute impaired glucoregulation in healthy adolescents [73]. It appears that regular pairings of sugar and caffeine through SBs may influence adiposity through some of these mechanisms. Indeed, a recent study based on 3 large cohorts found that drinking unsweetened coffee, may prevent weight gain, but this benefit was negated by adding sugar [74].

A key strength of our study lies in the dynamic DONALD cohort design, which enables repeated dietary and biosample collections from

the same individuals. The repeatedly measured metabolome in children uncovered potentially transient diet-metabolome associations, which may be missed in single point measurement. Our study applied multiple robust ML approaches, which generally yielded comparable selections; yet, the discrepancies also underscore the drawbacks of single-method reliance in high-dimensional data.

Further, the use of 3 standard adiposity measures, which assess the general and abdominal adiposity, enhances the translational utility and potential for our findings to be replicated by larger epidemiologic studies. To our knowledge, this study represents the first comprehensive exploration of the metabolome with SB and AS intakes and their associations with adiposity in young individuals. We demonstrate that our approach lends more insights, providing complimentary information on metabolic changes associated with intake, and their differences may reflect biologically meaningful processes.

Our study had some limitations such as potential measurement errors in self-reported dietary intakes. The interpretation of the associations of food-related metabolites and adiposity was limited by their concurrent measurements. Future studies may investigate the longitudinal associations of these metabolites with adiposity. We also acknowledge that, even with the repeated double cross-validation and bootstrap procedures, metabolite selection and the subsequent analysis were conducted on the same data set for maximum use, which could result in overly optimistic results in downstream regression analysis. The biochemical names of many metabolites in children samples could not be identified, limiting the comparison of our findings across age groups. Finally, the DONALD cohort's homogeneity and higher socioeconomic status than the general population [25] warrants cautious interpretation of the results. Nonetheless, its adiposity trends from birth to 14 y are comparable with 2 other German cohorts [75], thus our findings have reasonable generalizability.

In conclusion, we identified metabolomics signatures of SB and AS intake and their associations with anthropometric measures of adiposity in a well-characterized German birth cohort. If validated in other studies, these metabolomics profiles could further elucidate the underlying mechanisms through which these foods influence adiposity.

Acknowledgments

We thank the DONALD study participants and their families for their participation in this study. Figure 1 was created with BioRender.com. We thank Dr. rer. nat. Cédric Scherer and Yan Holtz for the R scripts (https://r-graph-gallery.com/circular-barplot.html) used in Figure 2.

Author contributions

The authors' responsibilities were as follows – UN, AF, MS, KO, SM: designed research; UA, IP, JG: conducted research; MM, JR, DA, AS, PK-R: provided essential materials; MES, SM: compiled dietary intake data; SM: analyzed data and wrote the first draft; KO: provided statistical supervision; SM, KO, UN: had primary responsibility for final content; and all authors: read and approved the final manuscript.

Conflict of interest

The authors report no conflicts of interests. Where authors are identified as personnel of the International Agency for Research on Cancer/WHO, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/WHO.

Funding

This study was funded by the German Research Foundation (DFG 406710821) and the Agence Nationale de la Recherche and partly supported by Diet–Body–Brain (DietBB), the Competence Cluster in Nutrition Research (Federal Ministry of Education and Research, FKZ:01EA1410A), and the PerMiCCion project (Federal Ministry of Education and Research, Grant 100554612). KO is supported by the German Research Foundation (DFG 460591722).

Data availability

The data described in the manuscript will be made available upon request, pending application and approval. Please submit your requests to Prof. Ute Nöthlings at epi@uni-bonn.de. The analytic code is available from the corresponding author.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ajcnut.2024.07.021.

References

[1] V. Calcaterra, H. Cena, V.C. Magenes, A. Vincenti, G. Comola, A. Beretta, et al., Sugar-sweetened beverages and metabolic risk in children and adolescents with obesity: a narrative review, Nutrients 15 (3) (2023) 702, https://doi.org/10.3390/nu15030702.
[2] T.G. Sánchez-Pimienta, C. Batis, C.K. Lutter, J.A. Rivera, Sugar-sweetened beverages are the main sources of added sugar intake in the Mexican population, J. Nutr. 146 (9) (2016) 1888S–1896S, https://doi.org/10.3945/jn.115.220301.
[3] J. Reedy, S.M. Krebs-Smith, Dietary sources of energy, solid fats, and added sugars among children and adolescents in the United States, J. Am. Diet, Assoc. 110 (10) (2010) 1477–1484, https://doi.org/10.1016/j.jada.2010.07.010.
[4] T. Andreyeva, Large state variation in sugar-sweetened beverage purchases: what we learn from the beverage industry data, Curr. Dev. Nutri. 5 (12) (2021) nzab128, https://doi.org/10.1093/cdn/nzab128.
[5] V.S. Malik, F.B. Hu, The role of sugar-sweetened beverages in the global epidemics of obesity and chronic diseases, Nat. Rev. Endocrinol. 18 (4) (2022) 205–218, https://doi.org/10.1038/s41574-021-00627-6.
[6] V.S. Malik, F.B. Hu, Sugar-sweetened beverages and cardiometabolic health: an update of the evidence, Nutrients 11 (8) (2019) 1840, https://doi.org/10.3390/nu11081840.
[7] P. Qin, Q. Li, Y. Zhao, Q. Chen, X. Sun, Y. Liu, et al., Sugar and artificially sweetened beverages and risk of obesity, type 2 diabetes mellitus, hypertension, and all-cause mortality: a dose–response meta-analysis of prospective cohort studies, Eur. J. Epidemiol. 35 (7) (2020) 655–671, https://doi.org/10.1007/s10654-020-00655-y.
[8] P.R. Trumbo, C.R. Rivers, Systematic review of the evidence for an association between sugar-sweetened beverage consumption and risk of obesity, Nutr. Rev. 72 (9) (2014) 566–574, https://doi.org/10.1111/nure.12128.
[9] F. Llaha, M. Gil-Lespinard, P. Unal, I. de Villasante, J. Castañeda, R. Zamora-Ros, Consumption of sweet beverages and cancer risk. A systematic review and meta-analysis of observational studies, Nutrients 13 (2) (2021) 516, https://doi.org/10.3390/nu13020516.
[10] A.F. Subar, V. Kipnis, R.P. Troiano, D. Midthune, D.A. Schoeller, S. Bingham, et al., Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study, Am. J. Epidemiol. 158 (1) (2003) 1–13, https://doi.org/10.1093/aje/kwg092.
[11] M.A. Knox, V.M. Oddo, L.P. Walkinshaw, J. Jones-Smith, Is the public sweet on sugary beverages? Social desirability bias and sweetened beverage taxes, Econ. Hum. Biol. 38 (2020) 100886, https://doi.org/10.1016/j.ehb.2020.100886.
[12] 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. 99 (6) (2014) 1286–1308, https://doi.org/10.3945/ajcn.113.076133.
[13] Y. Huang, C. Zheng, L.F. Tinker, M.L. Neuhouser, R.L. Prentice, Biomarker-based methods and study designs to calibrate dietary intake for assessing diet–disease associations, J. Nutr. 152 (3) (2022) 899–906, https://doi.org/10.1093/jn/nxab420.
[14] L.S. Freedman, V. Kipnis, D. Midthune, J. Commins, B. Barrett, V. Sagi-Kiss, et al., Establishing 24-hour urinary sucrose plus fructose as a predictive

biomarker for total sugars intake, Cancer Epidemiol. Biomarkers Prev. 31 (6) (2022) 1227–1232, https://doi.org/10.1158/1055-9965.EPI-21-1293.
[15] M.L. Neuhouser, R.L. Prentice, L.F. Tinker, J.W. Lampe, Enhancing capacity for food and nutrient intake assessment in Population Sciences Research, Annu. Rev. Public Health 44 (2023) 37–54, https://doi.org/10.1146/annurev-publhealth-071521-121621.
[16] N. Tasevska, D. Midthune, N. Potischman, A.F. Subar, A.J. Cross, S.A. Bingham, et al., Use of the predictive sugars biomarker to evaluate self-reported total sugars intake in the Observing Protein and Energy Nutrition (OPEN) study, Cancer Epidemiol. Biomarkers Prev. 20 (3) (2011) 490–500, https://doi.org/10.1158/1055-9965.EPI-10-0820.
[17] J.M. Beasley, M. Jung, N. Tasevska, W.W. Wong, A.M. Siega-Riz, D. Sotres-Alvarez, et al., Biomarker-predicted sugars intake compared with self-reported measures in US hispanics/latinos: results from the HCHS/sol solnas study, Public Health Nutr 19 (18) (2016) 3256–3264, https://doi.org/10.1017/S1368980016001580.
[18] S. Muli, J. Goerdten, K. Oluwagbemigun, A. Floegel, M. Schmid, U. Nöthlings, A systematic review of metabolomic biomarkers for the intake of sugar-sweetened and low-calorie sweetened beverages, Metabolites 11 (8) (2021) 546, https://doi.org/10.3390/metabo11080546.
[19] J.A. Rothwell, F. Madrid-Gambin, M. Garcia-Aloy, C. Andres-Lacueva, C. Logue, A.M. Gallagher, et al., Biomarkers of intake for coffee, tea, and sweetened beverages, Genes Nutr 13 (2018) 15, https://doi.org/10.1186/s12263-018-0607-5.
[20] R. Landberg, P. Karra, R. Hoobler, E. Loftfield, I. Huybrechts, J.I. Rattner, et al., Dietary biomarkers—an update on their validity and applicability in epidemiological studies, Nutr. Rev. (2023 Oct 3) nuad119, https://doi.org/10.1093/nutrit/nuad119.
[21] H. Gibbons, B.A. McNulty, A.P. Nugent, J. Walton, A. Flynn, M.J. Gibney, et al., Metabolomics approach to the identification of biomarkers of sugar-sweetened beverage intake, Am. J. Clin. Nutr. 101 (3) (2015) 471–477, https://doi.org/10.3945/ajcn.114.095604.
[22] T. Yan, L. Shi, K. Xu, J. Bai, R. Wen, X. Liao, et al., Habitual intakes of sugar-sweetened beverages associated with gut microbiota-related metabolites and metabolic health outcomes in young Chinese adults, Nutr. Metab. Cardiovasc. Dis. 33 (2) (2023) 359–368, https://doi.org/10.1016/j.numecd.2022.10.016.
[23] W. Perng, L. Tang, P.X. Song, M. Goran, M.M. Tellez Rojo, A. Cantoral, et al., Urate and nonanoate mark the relationship between sugar-sweetened beverage intake and blood pressure in adolescent girls: a metabolomics analysis in the element cohort, Metabolites 9 (5) (2019) 100, https://doi.org/10.3390/metabo9050100.
[24] B. Zhou, R. Ichikawa, L.D. Parnell, S.E. Noel, X. Zhang, S.N. Bhupathiraju, et al., Metabolomic links between sugar-sweetened beverage intake and obesity, J. Obes. 2020 (2020) 7154738, https://doi.org/10.1155/2020/7154738.
[25] I. Perrar, U. Alexy, U. Nöthlings, Cohort profile update-overview of over 35 years of research in the Dortmund Nutritional and Anthropometric Longitudinally Designed (DONALD) study, Eur. J. Nutr. 63 (3) (2024) 727–740, https://doi.org/10.1007/s00394-023-03290-x.
[26] W. Sichert-Hellert, M. Kersting, C. Chahda, R. Schäfer, A. Kroke, German food composition database for dietary evaluations in children and adolescents, J. Food Compos. Anal. 20 (1) (2007) 63–70, https://doi.org/10.1016/j.jfca.2006.05.004.
[27] J.H. Cummings, A.M. Stephen, Carbohydrate terminology and classification, Eur. J. Clin. Nutr. 61 (Suppl 1) (2007) S5–S18, https://doi.org/10.1038/sj.ejcn.1602936.
[28] A.S. Rosario, B.M. Kurth, H. Stolzenberg, U. Ellert, H. Neuhauser, Body mass index percentiles for children and adolescents in Germany based on a nationally representative sample (KiGGS 2003–2006), Eur. J. Clin. Nutr. 64 (4) (2010) 341–349, https://doi.org/10.1038/ejcn.2010.8.
[29] P. Deurenberg, J.J. Pieters, J.G. Hautvast, The assessment of the body fat percentage by skinfold thickness measurements in childhood and young adolescence, Br. J. Nutr. 63 (2) (1990) 293–303, https://doi.org/10.1079/BJN19900116.
[30] M.L. Booth, A.D. Okely, T.N. Chey, A. Bauman, The reliability and validity of the Adolescent Physical Activity Recall Questionnaire, Med. Sci. Sports Exerc. 34 (12) (2002) 1986–1995, https://doi.org/10.1097/00005768-200212000-00019.
[31] A.M. Evans, C.D. DeHaven, T. Barrett, M. Mitchell, E. Milgram, Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems, Anal. Chem. 81 (16) (2009) 6656–6667, https://doi.org/10.1021/ac901536h.
[32] N. Bararpour, F. Gilardi, C. Carmeli, J. Sidibe, J. Ivanisevic, T. Caputo, et al., DBnorm as an R package for the comparison and selection of appropriate statistical methods for batch effect correction in metabolomic studies, Sci. Rep. 11 (1) (2021) 5657, https://doi.org/10.1038/s41598-021-84824-3.
[33] L. Shi, J.A. Westerhuis, J. Rosén, R. Landberg, C. Brunius, Variable selection and validation in multivariate modelling, Bioinformatics 35 (6) (2019) 972–980, https://doi.org/10.1093/bioinformatics/bty710.
[34] J. Liang, C. Wang, D. Zhang, Y. Xie, Y. Zeng, T. Li, et al., VSOLassoBag: a variable-selection oriented lasso bagging algorithm for biomarker discovery in omic-based translational research, J. Genet. Genomics. 50 (3) (2023) 151–162, https://doi.org/10.1016/j.jgg.2022.12.005.
[35] J.H. Kim, Multicollinearity and misleading statistical results, Korean J. Anesthesiol. 72 (6) (2019) 558–569, https://doi.org/10.4097/kja.19087.
[36] A.J. Sankoh, M.F. Huque, S.D. Dubey, Some comments on frequently used multiple endpoint adjustment methods in clinical trials, Stat. Med. 16 (22) (1997) 2529–2542, https://doi.org/10.1002/(SICI)1097-0258(19971130)16:22<2529::AID-SIM692>3.0.CO;2-J.
[37] R. Wei, J. Wang, M. Su, E. Jia, S. Chen, T. Chen, et al., Missing value imputation approach for mass spectrometry-based metabolomics data, Sci. Rep. 8 (1) (2018) 663, https://doi.org/10.1038/s41598-017-19120-0.
[38] M.D. Wilson, M.D. Ponzini, S.L. Taylor, K. Kim, Imputation of missing values for multi-biospecimen metabolomics studies: bias and effects on statistical validity, Metabolites 12 (7) (2022) 671, https://doi.org/10.3390/metabo12070671.
[39] M. Kokla, J. Virtanen, M. Kolehmainen, J. Paananen, K. Hanhineva, Random forest-based imputation outperforms other methods for imputing LC-MS metabolomics data: a comparative study, BMC Bioinformatics 20 (1) (2019) 492, https://doi.org/10.1186/s12859-019-3110-0.
[40] M.J. Lewis, A. Spiliopoulou, K. Goldmann, C. Pitzalis, P. McKeigue, M.R. Barnes, nestedcv: an R package for fast implementation of nested cross-validation with embedded feature selection designed for transcriptomics and high-dimensional data, Bioinform. Adv. 3 (1) (2023) vbad048, https://doi.org/10.1093/bioadv/vbad048.
[41] B.M. Popkin, C. Hawkes, Sweetening of the global diet, particularly beverages: patterns, trends, and policy responses, Lancet Diabetes Endocrinol 4 (2) (2016) 174–186, https://doi.org/10.1016/S2213-8587(15)00419-2.
[42] B.A. Magnuson, M.C. Carakostas, N.H. Moore, S.P. Poulos, A.G. Renwick, Biological fate of low-calorie sweeteners, Nutr. Rev. 74 (11) (2016) 670–689, https://doi.org/10.1093/nutrit/nuw032.
[43] A.R. Basson, A. Rodriguez-Palacios, F. Cominelli, Artificial sweeteners: history and new concepts on inflammation, Front. Nutr. 8 (2021) 746247, https://doi.org/10.3389/fnut.2021.746247.
[44] C.C. Cohen, D. Dabelea, G. Michelotti, L. Tang, K. Shankar, M.I. Goran, et al., Metabolome alterations linking sugar-sweetened beverage intake with dyslipidemia in youth: the Exploring Perinatal Outcomes Among Children (EPOCH) study, Metabolites 12 (6) (2022) 559, https://doi.org/10.3390/metabo12060559.
[45] N. Tasevska, S.A. Runswick, A. McTaggart, S.A. Bingham, Urinary sucrose and fructose as biomarkers for sugar consumption, Cancer Epidemiol. Biomarkers Prev. 14 (5) (2005) 1287–1294, https://doi.org/10.1158/1055-9965.EPI-04-0827.
[46] I. Perrar, N. Gray, G.G. Kuhnle, T. Remer, A.E. Buyken, U. Alexy, Sugar intake among German adolescents: trends from 1990 to 2016 based on biomarker excretion in 24-h urine samples, Br. J. Nutr. 124 (2) (2020) 164–172, https://doi.org/10.1017/S0007114520000665.
[47] J.C. Louie, Objective biomarkers for total added sugar intake—are we on a wild goose chase? Adv. Nutr. 11 (6) (2020) 1429–1436, https://doi.org/10.1093/advances/nmaa093.
[48] M. Dambrova, M. Makrecka-Kuka, J. Kuka, R. Vilskersts, D. Nordberg, M.M. Attwood, et al., Acylcarnitines: nomenclature, biomarkers, therapeutic potential, drug targets, and clinical trials, Pharmacol. Rev. 74 (3) (2022) 506–551, https://doi.org/10.1124/pharmrev.121.000408.
[49] R. Zhang, K. Zhang, Mitochondrial NAD kinase in health and disease, Redox. Biol. 60 (2023) 102613, https://doi.org/10.1016/j.redox.2023.102613.
[50] D.M. Zheng, Z.N. An, M.H. Ge, D.Z. Wei, D.W. Jiang, X.J. Xing, et al., Medium & long-chain acylcarnitine's relation to lipid metabolism as potential predictors for diabetic cardiomyopathy: a metabolomic study, Lipids Health Dis 20 (1) (2021) 151, https://doi.org/10.1186/s12944-021-01576-9.
[51] X. Liu, P. Zheng, X. Zhao, Y. Zhang, C. Hu, J. Li, et al., Discovery and validation of plasma biomarkers for major depressive disorder classification based on liquid chromatography–mass spectrometry, J. Proteome Res. 14 (5) (2015) 2322–2330, https://doi.org/10.1021/acs.jproteome.5b00144.
[52] R. Wedekind, A. Kiss, P. Keski-Rahkonen, V. Viallon, J.A. Rothwell, A.J. Cross, et al., A metabolomic study of red and processed meat intake and acylcarnitine concentrations in human urine and blood, Am. J. Clin. Nutr. 112 (2) (2020) 381–388, https://doi.org/10.1093/ajcn/nqaa140.
[53] L.V. Griebsch, E.L. Theiss, D. Janitschke, V.K.J. Erhardt, T. Erhardt, E.C. Haas, et al., Aspartame and its metabolites cause oxidative stress and mitochondrial and lipid alterations in SH-SY5Y cells, Nutrients 15 (6) (2023) 1467, https://doi.org/10.3390/nu15061467.

[54] K. Tamae, K. Kawai, S. Yamasaki, K. Kawanami, M. Ikeda, K. Takahashi, et al., Effect of age, smoking and other lifestyle factors on urinary 7-methylguanine and 8-hydroxydeoxyguanosine, Cancer Sci 100 (4) (2009) 715–721, https://doi.org/10.1111/j.1349-7006.2009.01088.x.
[55] D.D.B. Holthuijsen, M.J.L. Bours, E.H.V. Roekel, S.O. Breukink, M.L.G. Janssen-Heijnen, E.T.P. Keulen, et al., Longitudinal associations of adherence to the Dietary World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR) and Dutch Healthy Diet (DHD) recommendations with plasma kynurenines in colorectal cancer survivors after treatment, Nutrients 14 (23) (2022) 5151, https://doi.org/10.3390/nu14235151.
[56] H.M. Francis, R.J. Stevenson, L.S. Tan, L. Ehrenfeld, S. Byeon, T. Attuquayefio, et al., Kynurenic acid as a biochemical factor underlying the association between western-style diet and depression: a cross-sectional study, Front. Nutr. 9 (2022) 945538, https://doi.org/10.3389/fnut.2022.945538.
[57] B. Pineda, V. Campos-Peña, R. Lugo-Huitrón, C. Ríos, V. Pérez-de la Cruz, The kynurenine pathway at the interface between neuroinflammation, oxidative stress, and neurochemical disturbances: emphasis in schizophrenia, in: A. Dietrich-Muszalska, V. Chauhan, S. Grignon (Eds.), Studies on psychiatric disorders, Springer, New York, NY, United States, 2015, pp. 245–268, https://doi.org/10.1007/978-1-4939-0440-2_13.
[58] L. Xu, A.J. Sinclair, M. Faiza, D. Li, X. Han, H. Yin, et al., Furan fatty acids – beneficial or harmful to health? Prog Lipid Res 68 (2017) 119–137, https://doi.org/10.1016/j.plipres.2017.10.002.
[59] Z. Miao, F.F. Zeng, Y. Tian, C. Xiao, Y. Yan, Z. Jiang, et al., Furan fatty acid metabolite CMPF is associated with lower risk of type 2 diabetes, but not chronic kidney disease: a longitudinal population-based cohort study, Am. J. Clin. Nutr. 118 (3) (2023) 637–645, https://doi.org/10.1016/j.ajcnut.2023.07.016.
[60] L. Dai, Z.A. Massy, P. Stenvinkel, N.C. Chesnaye, I.A. Larabi, J.C. Alvarez, et al., The association between TMAO, CMPF, and clinical outcomes in advanced chronic kidney disease: results from the European Quality (EQUAL) study, Am. J. Clin. Nutr. 116 (6) (2022) 1842–1851, https://doi.org/10.1093/ajcn/nqac278.
[61] K.D. Laugero, N.L. Keim, A diet pattern characterized by sugar-sweetened beverages is associated with lower decision-making performance in the Iowa gambling task, elevated stress exposure, and altered autonomic nervous system reactivity in men and women, Nutrients 15 (18) (2023) 3930, https://doi.org/10.3390/nu15183930.
[62] L. Libuda, U. Alexy, A.E. Buyken, W. Sichert-Hellert, P. Stehle, M. Kersting, Consumption of sugar-sweetened beverages and its association with nutrient intakes and diet quality in German children and adolescents, Br. J. Nutr. 101 (10) (2009) 1549–1557, https://doi.org/10.1017/S0007114508094671.
[63] M. Ulaszewska, N. Vázquez-Manjarrez, M. Garcia-Aloy, R. Llorach, F. Mattivi, L.O. Dragsted, et al., Food intake biomarkers for apple, pear, and stone fruit, Genes Nutr 13 (2018) 29, https://doi.org/10.1186/s12263-018-0620-8.
[64] H. Reiland, J. Slavin, Systematic review of pears and health, Nutr. Today. 50 (6) (2015) 301–305, https://doi.org/10.1097/NT.0000000000000112.
[65] K. Elfhag, P. Tynelius, F. Rasmussen, Sugar-sweetened and artificially sweetened soft drinks in association to restrained, external and emotional eating, Physiol. Behav. 91 (2–3) (2007) 191–195, https://doi.org/10.1016/j.physbeh.2007.02.005.
[66] E.S. Hwang, S.B. Song, Possible adverse effects of high-dose nicotinamide: mechanisms and safety assessment, Biomolecules 10 (5) (2020) 687, https://doi.org/10.3390/biom10050687.
[67] S. Mayengbam, H. Virtanen, D.S. Hittel, C. Elliott, R.A. Reimer, H.J. Vogel, et al., Metabolic consequences of discretionary fortified beverage consumption containing excessive vitamin B levels in adolescents, PLoS One 14 (1) (2019) e0209913, https://doi.org/10.1371/journal.pone.0209913.
[68] W.L. Lowe Jr., J.R. Bain, M. Nodzenski, A.C. Reisetter, M.J. Muehlbauer, R.D. Stevens, et al., Maternal BMI and glycemia impact the fetal metabolome, Diabetes Care 40 (7) (2017) 902–910, https://doi.org/10.2337/dc16-2452.
[69] S. Yan, S. Chen, Y. Liu, H. Liang, X. Zhang, Q. Zhang, et al., Associations of serum carotenoids with visceral adiposity index and lipid accumulation product: a cross-sectional study based on NHANES 2001-2006, Lipids Health Dis 22 (1) (2023) 209, https://doi.org/10.1186/s12944-023-01945-6.
[70] J.L. Temple, Caffeine use in children: what we know, what we have left to learn, and why we should worry, Neurosci. Biobehav. Rev. 33 (6) (2009) 793–806, https://doi.org/10.1016/j.neubiorev.2009.01.001.
[71] A.C. Sylvetsky, E.F. Blake, A.J. Visek, S. Halberg, K. Comstock, K.D. Essel, et al., Feasibility and acceptability of a randomized controlled trial to investigate withdrawal symptoms in response to caffeinated sugary drink cessation among children, Contemp. Clin. Trials Commun. 22 (2021) 100791, https://doi.org/10.1016/j.conctc.2021.100791.
[72] L.L. Moisey, L.E. Robinson, T.E. Graham, Consumption of caffeinated coffee and a high carbohydrate meal affects postprandial metabolism of a subsequent oral glucose tolerance test in young, healthy males, Br. J. Nutr. 103 (6) (2010) 833–841, https://doi.org/10.1017/S0007114509992406.
[73] J. Shearer, R.A. Reimer, D.S. Hittel, M.A. Gault, H.J. Vogel, M.S. Klein, Caffeine-containing energy shots cause acute impaired glucoregulation in adolescents, Nutrients 12 (12) (2020) 3850, https://doi.org/10.3390/nu12123850.
[74] M. Henn, A.J. Glenn, W.C. Willett, M.A. Martínez-González, Q. Sun, F.B. Hu, Changes in coffee intake, added sugar and long-term weight gain—results from three large prospective US cohort studies, Am. J. Clin. Nutr. 118 (6) (2023) 1164–1171, https://doi.org/10.1016/j.ajcnut.2023.09.023.
[75] R. von Kries, M.J. Müller, J. Heinrich, Early prevention of childhood obesity: another promise or a reliable path for battling childhood obesity? Obes. Facts. 7 (2) (2014) 77–81, https://doi.org/10.1159/000362190.

3.3 Publication 3

Identification and replication of urine metabolites associated with short-term and habitual intake of sweet and fatty snacks in European children and adolescents. https://doi.org/10.1016/j.tjnut.2024.09.026

The Journal of Nutrition 154 (2024) 3274–3285

American Society for Nutrition
Excellence in Nutrition Research and Practice

JN THE JOURNAL OF NUTRITION

journal homepage: https://jn.nutrition.org/

Genomics, Proteomics, and Metabolomics

Identification and Replication of Urine Metabolites Associated With Short-Term and Habitual Intake of Sweet and Fatty Snacks in European Children and Adolescents

Jantje Goerdten [1,‡,*], Samuel Muli [2], Jodi Rattner [3], Mira Merdas [3], David Achaintre [3], Li Yuan [1,‡], Stefaan De Henauw [4,‡], Ronja Foraita [1,‡], Monica Hunsberger [5,‡], Inge Huybrechts [3], Lauren Lissner [5,‡], Dénes Molnár [6,‡], Luis A Moreno [7,8,‡], Paola Russo [9,‡], Toomas Veidebaum [10,‡], Krasimira Aleksandrova [1,11,‡], Ute Nöthlings [2], Kolade Oluwagbemigun [2], Pekka Keski-Rahkonen [3,†], Anna Floegel [1,12,†,‡]

[1] *Leibniz Institute for Prevention Research and Epidemiology (BIPS), Bremen, Germany;* [2] *Unit of Nutritional Epidemiology, Department of Nutrition and Food Sciences, University of Bonn, Bonn, Germany;* [3] *International Agency for Research on Cancer (IARC), Lyon, France;* [4] *Department of Public Health and Primary Care, Ghent University, Ghent, Belgium;* [5] *School of Public Health and Community Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden;* [6] *Department of Pediatrics, Medical School, University of Pécs, Pécs, Hungary;* [7] *GENUD (Growth, Exercise, NUtrition and Development) Research Group, Faculty of Health Sciences, University of Zaragoza, Instituto Agroalimentario de Aragón (IA2) and Instituto de Investigación Sanitaria Aragón (IIS Aragón), Zaragoza, Spain;* [8] *Consorcio CIBER, M.P. Fisiopatología de la Obesidad y Nutrición (CIBERObn), Instituto de Salud Carlos III (ISCIII), Madrid, Spain;* [9] *Institute of Food Sciences, CNR, Avellino, Italy;* [10] *National Institute for Health Development, Tallinn, Estonia;* [11] *Faculty of Human and Health Sciences, University of Bremen, Bremen, Germany;* [12] *Section of Dietetics, Faculty of Agriculture and Food Sciences, Hochschule Neubrandenburg—University of Applied Sciences, Neubrandenburg, Germany*

A B S T R A C T

Background: Intake of sweet and fatty snacks may partly contribute to the occurrence of obesity and other health conditions in childhood. Traditional dietary assessment methods may be limited in accurately assessing the intake of sweet and fatty snacks in children. Metabolite biomarkers may aid the objective assessment of children's food intake and support establishing diet–disease relationships.
Objectives: The present study aimed to identify biomarkers of sweet and fatty snack intake in 2 independent cohorts of European children.
Methods: We used data from the IDEFICS/I.Family cohort from baseline (2007/2008) and 2 follow-up examination waves (2009/2010 and 2013/2014). In total, 1788 urine samples from 599 children were analyzed for untargeted metabolomics using high-resolution liquid chromatography-mass spectrometry. Short-term dietary intake was assessed by 24-h dietary recalls, and habitual dietary intake was calculated with the National Cancer Institute method. Data from the Dortmund Nutritional and Anthropometric Longitudinal Designed (DONALD) cohort of 24-h urine samples ($n = 567$) and 3-d weighted dietary records were used for external replication of results. Multivariate modeling with unbiased variable selection in R algorithms and linear mixed models were used to identify novel biomarkers. Metabolite features significantly associated with dietary intake were then annotated.
Results: In total, 66 metabolites were discovered and found to be statistically significant for chocolate candy; cakes, puddings, and cookies; candy and sweets; ice cream; and crisps. Most of the features ($n = 62$) could not be annotated. Short-term and habitual chocolate intake were positively associated with theobromine, xanthosine, and cyclo(L-prolyl-L-valyl). These results were replicated in the DONALD cohort. Short-term candy and sweet intake was negatively associated with octenoylcarnitine.

Abbreviations: BFI, biomarkers of food intake; DONALD, Dortmund Nutritional and Anthropometric Longitudinal Designed; FFQ, food frequency questionnaire; IARC, International Agency for Research on Cancer; IDEFICS, Identification and Prevention of Dietary- and Lifestyle-induced Health Effects in Children and Infants; MUVR, multivariate modeling with minimally biased variable selection in R; NCI, United States National Cancer Institute; PLS, partial least squares; RF, random forest; 24-HDR, 24-h dietary recall; 3d-WDR, 3-d weighted dietary record.

* Corresponding author. *E-mail addresses:* goerdten@leibniz-bips.de, sec-epi@leibniz-bips.de (J. Goerdten).
† PK-R and AF contributed equally to this work.
‡ on behalf of the IDEFICS/I.Family consortia

https://doi.org/10.1016/j.tjnut.2024.09.026
Received 26 April 2024; Received in revised form 19 September 2024; Accepted 23 September 2024; Available online 25 September 2024

Conclusions: Of the potential metabolite biomarkers of sweet and fatty snacks in children, 3 biomarkers of chocolate intake, namely theobromine, xanthosine, and cyclo(L-prolyl-L-valyl), are externally replicated. However, these potential biomarkers require further validation in children.

Keywords: biomarker identification, metabolite biomarkers of food intake, untargeted metabolomics, sweet and fatty snacks, children and adolescents

Introduction

In epidemiologic studies, dietary intake is commonly assessed through self-reports using instruments such as food frequency questionnaires (FFQs), 24-h dietary recalls (24-HDRs), and 3-d weighted dietary records (3d-WDRs) [1]. These dietary assessment methods present some challenges, such as the potential for misreporting [2]. Although progress has been made in generating evidence for diet–disease relationships in adults, evidence is lacking for children and adolescents [3]. This might be partially due to the additional difficulties in dietary intake assessment for children and adolescents, namely, unstructured eating patterns, concerns with self-image, or problems in conceptualizing time, to mention only a few [4]. Hence, objective measures of nutrient and dietary intake are needed to aid or even replace the traditional assessment methods. Indeed, dietary biomarkers, such as urinary nitrogen [5] or vitamin C [6], have been used as objective indicators of dietary intake in dietary validation studies [7,8].

The rise of omics technologies has paved the way for objective measures of dietary intake, also known as biomarkers of food intake (BFI) [1,9,10]. Specifically, metabolomics has emerged as a powerful discovery method for the analysis of biospecimens, such as urine, allowing for the subsequent identification of novel BFIs [11]. In untargeted metabolomics, low molecular weight compounds, called metabolites, are comprehensively analyzed without previous selection of targeted analytes. Metabolites are the end products of metabolism, many of which reflect short-term or habitual food intakes [10]. Furthermore, measuring the food metabolome through untargeted metabolomics can yield a vast number of metabolites [10]. Urine biosamples are a great resource because the collection is noninvasive and can be collected at home by the study participants [12]. Therefore, analyzing the urine-based food metabolome offers a unique opportunity to identify novel BFIs.

Over the last decade, there have been enormous efforts to identify biomarkers of various nutrients, foods, food groups, or dietary patterns [13–23]. However, most of the studies have focused on adult populations, with only a few including children and/or adolescents [13,24,25]. Furthermore, most studies only internally validated their findings and lack external replication [26]. External validation is an important step in increasing the reliability of BFI.

For children and adolescents, the most studied BFIs are for fruit and vegetable intake [25,27]. Additionally, 1 study identified BFIs for meat and fish intake [24]. To date, to our knowledge, there are no candidate BFIs for sweet and fatty snacks in children and adolescents. This gap exists partly because few biomarker discovery studies focus on children and partly because sweet and fatty snack intake is especially difficult to assess [4]. Moreover, only a few candidate BFIs exist for adult populations, mainly for cocoa, liquorice, and potato crisps [17,23].

BFIs for sweet and fatty snacks in children are missing. If identified, these BFIs may provide important insights into understanding childhood diet–disease associations. Hence, this study aimed to identify novel BFIs of short-term and habitual intake of sweet and fatty snack foods in the repeatedly measured food metabolome from 2 independent longitudinal cohorts of European children and adolescents.

Methods

IDEFICS/I.Family

The Identification and Prevention of Dietary- and Lifestyle-induced Health Effects in Children and Infants (IDEFICS) and I.Family cohort served as the main cohort for identifying biomarkers [28,29]. Data were gathered in 8 study centers across Europe—Belgium, Cyprus, Estonia, Germany, Hungary, Italy, Spain, and Sweden [29]. The study aimed to determine the etiology of overweight, obesity, and related health outcomes. Thus, repeated measurements of lifestyle, behavior, and medical parameters were taken, at baseline and over 5 follow-up time points. Children were recruited from kindergarten and school settings [28]. The baseline examination, conducted between 2007 and 2008, included >16,000 children aged 2 to10 y [29]. Ethical approval was obtained from the corresponding national or local ethics committees of the participating countries. A more in-depth description of the IDEFICS/I.Family cohort can be found elsewhere [28,29].

For this study, a random subsample ($N = 600$) of the original cohort was selected. Sample size calculations were performed a priori based on metabolite measurements at 1 time point. We expected a statistical power of >80%. Eligible participants had available urine samples, dietary, demographic, and anthropometric data at baseline and 2 follow-up time points. Figure 1 shows flow diagram of the study sample selection. Data from 3 time points, including baseline, were used in this study. These time points are referred to as baseline (W1), second examination wave (W2), and third wave (W3) hereafter. W2 data were collected from 2009 to 2010, and data collection for W3 took place between 2013 and 2014.

Covariate assessment

Anthropometric data from all participants were collected using standardized methods [28–30]. Body weight of the study participants was measured by a trained study nurse using a TANITA 418/420 digital scale [31]. Height was measured using a Seca 225/213 stadiometer [31]. BMI was calculated by dividing the weight in kilograms by the squared height in metres. Age-specific and sex-specific BMI *z*-scores according to Cole and Lobstein [32] were used in the statistical analysis. Demographic data, such as age, sex, and country of residence were routinely collected during baseline and follow-ups [28,29].

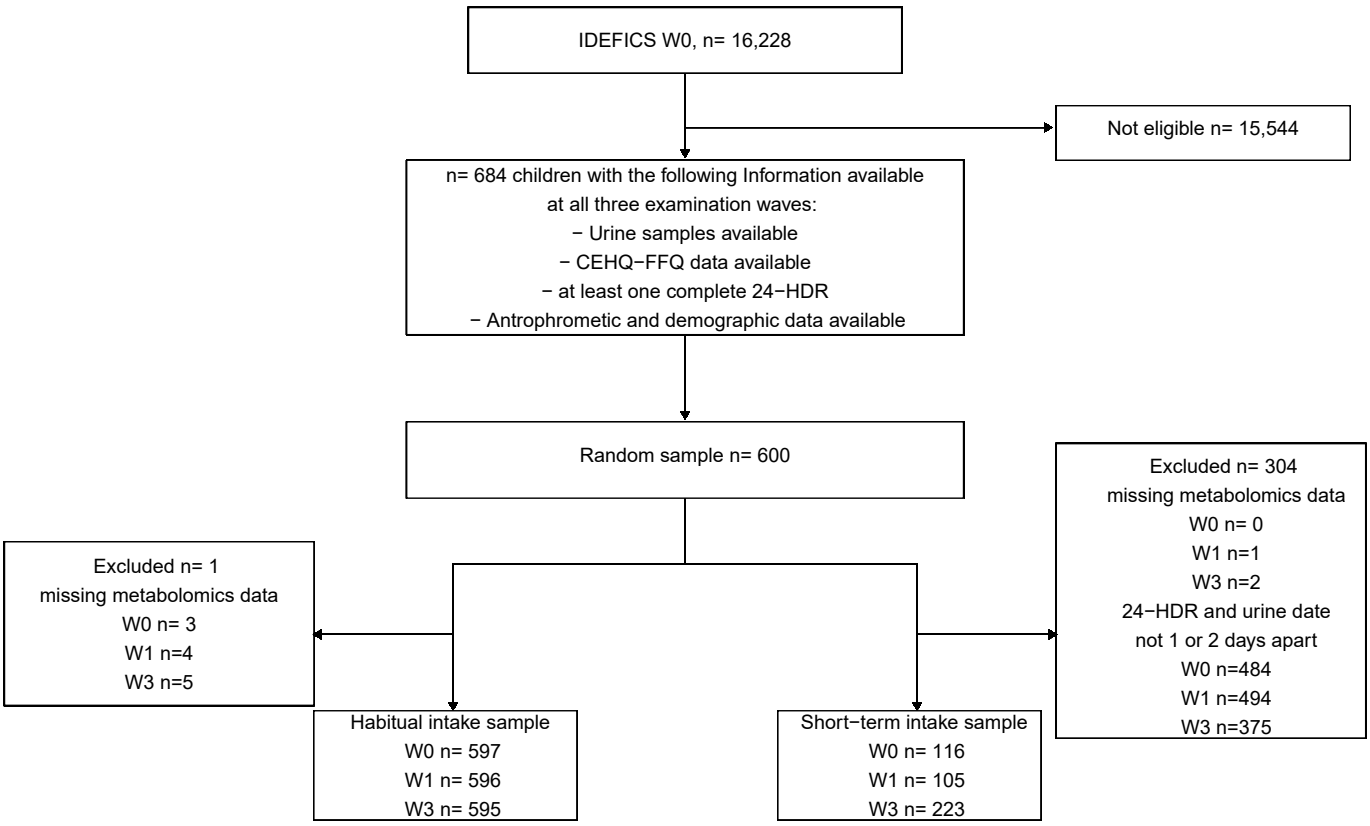

FIGURE 1. Flow diagram of the study sample selection in the discovery cohort (IDEFICS/I.Family). Abbreviations: CEHQ, Children's Eating Habits Questionnaire; FFQ, food frequency questionnaire; 24-HDR, 24-h dietary recall.

Dietary assessment

The dietary intake of participants was recorded using the 24-HDR and FFQ instruments. For the 24-HDR, a computer-based program, the Self-Administered Children and Infant Nutrition Assessment, was used at W1 and W2, and a web-based program, the Self-Administered Children, Adolescents, and Adult Nutrition Assessment, was used at W3. The FFQ section of the Children's Eating Habits Questionnaire was applied at each follow-up time point [33,34].

Study participants were asked to complete ≥2 weekday 24-HDRs and 1 on the weekend. If these were not completed, participants were reminded by telephone or e-mail [35]. The 24-HDR recorded the dietary intake (ie, the type and amount of foods and beverages) over the last 24 h [36]. The intake was structured into 6 meal occasions: breakfast (first intake after waking up), mid-morning snack, lunch, afternoon snack, dinner, and evening snack, with the possibility to add more eating occasions if required. Quantities of foods were estimated using standardized photographs of serving sizes, standard portions, customary packaging sizes, and meals in pieces or slices. For the 24-HDR based on the Self-Administered Children and Infant Nutrition Assessment program, a proxy respondent, that is, primary caretaker of the child, recalled the dietary intakes. Otherwise, children or adolescents recalled the dietary intake of the last 24 h themselves, with the help from primary caretakers if needed. A dietician or trained study nurse was present during each 24-HDR to answer any questions [30]. Several studies were undertaken to validate the 24-HDR in the IDEFICS/I.Family cohort [33].

The FFQ comprised 60 food items grouped into 15 food groups: specifically, vegetables, fruit, breakfast cereals, milk, yogurt, cheese, fish, meat and meat products, eggs, meat replacement products and soy products, spreadable products, cereal products and snacks, plant oil, and beverages [37]. Possible answers for the FFQ section were never/less than once a week, 1–3 times a week, 4–6 times a week, 1 time a day, 2 times a day, and 3 times or more a day. The previous month was set as the reference period for the FFQ. During W1 and W2, specifically, the primary caretaker of the child responded on behalf of the child. However, in W3, adolescents aged 12 y and over completed the FFQ section themselves. The measured food consumption of the FFQ was previously validated against nutrients measured in blood and urine [35,38].

Short-term dietary intake

Short-term dietary intake was defined as any food intake the day or 2 d before the urine collection. To derive the short-term dietary intake of chocolate candy (including cocoa powder used for chocolate beverages); chocolate and nut spread; potato crisps; jelly candy; candy and sweets (nonchocolate); cakes, puddings, and cookies; and ice cream, the dietary intakes of these foods were estimated in grams from the 24-HDR data. If several foods could be classified as the defined food group, the intake amounts were summed up. An overview of the food names used from the 24-HDR to derive the food group intakes is provided in the Supplemental Material. For some children, the difference between urine sample collection and assessed dietary intake was >2 d. These children were excluded from the analysis

of short-term food intake. Children for whom no metabolomics data were available owing to laboratory error were also excluded. Thus, the final study sample for short-term dietary intake was as follows: $n = 444$—for W1: $n = 116$, W2: $n = 105$, and W3: $n = 223$. There were 45 participants with 3 repeated measurements, 58 had 2, and 193 had 1 measurement.

We performed 2 sensitivity analyses. In the first, the short-term intake sample was limited to study participants with urine samples collected on the same day of the 24-HDR. In the second, we included coffee and tea intake as a further covariate in the statistical model to investigate any changes in the results for chocolate candy intake. Coffee and tea intake was derived from the 24-HDR (Supplemental Material).

Habitual dietary intake

Habitual dietary intake was defined as the estimated usual dietary intake per day for each participant. The method developed by the United States National Cancer Institute (NCI) was used to estimate individual habitual dietary intake (grams per day) [39]. To calculate habitual dietary intake for the study population the food groups defined from the 24-HDR and the dietary frequency questions from the FFQ were grouped (Supplemental Material) and analyzed according to Kipnis et al. [40]. 24-HDRs with an estimated total energy intake below 500 kcal were excluded from the analysis. The computation of the habitual dietary intake was stratified by sex and adjusted for age and BMI *z*-score. The habitual dietary intake was calculated for chocolate candy (including cocoa powder); savory and fatty snacks; candy and sweets (nonchocolate); crisps (potato and other); cakes, puddings and cookies; and ice cream. Metabolomics data were unavailable for very few children owing to laboratory errors. Thus, the final study sample consisted of $n = 1788$—for W1, $n = 597$; for W2, $n = 596$; and for W3, $n = 595$, respectively. All children who were included in the study sample for short-term dietary intake were also included in the study sample for habitual dietary intake.

Urine sample collection

Sample collection for the IDEFICS/I.Family cohort followed a standardized procedure across all participating study centers [41]. The morning urine was collected at home by the primary caregivers directly after waking. The caregivers received a urine collection kit and an instruction sheet [41]. Caregivers were asked to cool the urine sample in the refrigerator if the time until arrival at the study center was longer than 2 h. At the study center, the urine samples were cooled to −20 °C. At regular intervals, the biosamples were shipped on dry ice to the central biorepository, where they were frozen at −20 °C and stored [35,42].

Dortmund Nutritional and Anthropometric Longitudinal Designed

For the external replication of the identified metabolites the Dortmund Nutritional and Anthropometric Longitudinal Designed (DONALD) cohort was used [43]. The DONALD cohort is an open cohort study with 2375 participants recruited between 1985 and 2022 [44]. The participants were followed from infancy (3 mo of age) to adulthood at regular intervals during which dietary, health, and developmental data were collected. Dietary data were collected with 3d-WDR at each examination interval. Parents or children were asked to weigh their food and fill in the dietary records. Dieticians checked the records for plausibility. 24-h urine samples were collected at every interval from age 3 y onward. The urine samples were collected by the parent at home on the third day of the 3d-WDR. The biological samples were stored at the DONALD study center. Details of the DONALD cohort are available elsewhere [43,44].

For this project, a random sample of 300 children with ≥2 available urine samples and 2 3d-WDR were selected. In total, 600 urine samples were available. The final DONALD replication sample included 297 participants, of whom 270 had repeated measurements and 27 had single measurements available. The reduction in participants was due to an incomplete 3d-WDR and/or a mismatch between the 3d-WDR date and the urine collection date. This resulted in 567 measurements in the replication cohort.

Sample preparation and randomization

For this study, 1800 urine samples from the IDEFICS/I.Family cohort and 600 urine samples from the DONALD cohort were initially shipped from Bremen and Bonn, both in Germany, to the laboratory at the International Agency for Research on Cancer (IARC) for metabolomics analyses. The study samples were anonymized and randomized before shipment. The repeated samples for each study participant were analyzed next to each other in random order, and sample pairs were randomized across the batch. Randomization of the samples was further stratified by country for the multicenter IDEFICS/I.Family cohort samples, to ensure an equal proportion of samples from each country on each plate. Quality control samples were prepared from a sample pool, which was created by mixing small aliquots of all urine samples. Blank samples were also prepared in the same way as the urine samples, with only the urine being omitted in the process. Each 96-well plate included 4 individually prepared quality control samples and 2 blanks. Further details on the sample preparation can be found in the Supplemental Material.

Laboratory analysis and preprocessing

Study samples were analyzed using a Q Exactive mass spectrometer with heated electrospray ionization (HESI-II) coupled to a Dionex UltiMate 3000 Binary UHPLC system (ThermoFisher Scientific Inc.). Ten independent analytical batches consisting of 2 individual 96-well plates were analyzed. The mass spectrometer was operated in polarity-switching electrospray ionization mode (positive and negative ionization mode) to expand the coverage of the metabolome.

Preprocessing of the raw data was performed using Compound Discoverer 3.3 software (ThermoFischer Scientific). Metabolomic feature alignment between samples was performed with a maximum retention time window of 0.05 min and a mass tolerance of 5 ppm. Metabolite features were put forward into the feature table only if they were present in ≥2% of the overall samples. Metabolite features present in every blank sample were excluded, unless 5-fold greater in average intensity in samples. Furthermore, metabolite features absent in >95% of the study population were removed, and metabolite features that were missing in ≥30% of consumers (participants who habitually/acutely ate a food) were excluded. For the remaining metabolite features, missing values were imputed with half of the minimum value, that is, intensities, in each analytical batch, assuming

missingness due to limits of detection [45]. See the Supplemental Material for more information on the data preprocessing.

Statistical analyses

All analyses were performed using R 4.2.2 [46], primarily with the following packages: tidyverse (version 1.3.2) [47], nlme (version 3.1-160) [48], and multivariate modeling with minimally biased variable selection in R (MUVR; version 0.0.975) [49]. Before statistical analyses, all metabolite feature variables were log-transformed and *z*-standardized.

MUVR algorithm was used to identify the most predictive metabolite features for the dietary intakes (short-term or habitual) [50]. This algorithm incorporates recursive variable selection within a repeated double cross-validation scheme. MUVR offers approaches for feature selection in the presence of a large number of variables, namely partial least squares (PLSs) and random forest (RF). Both modeling approaches can be performed as either a regression or a classification analysis. Furthermore, MUVR can handle repeated measurements over time; this can be included in the algorithms by providing the subject ID.

For the short-term intake analysis, a classification analysis was performed for which the intake variables were dichotomized (0 for nonconsumers with intake = 0 g and 1 for consumers with intake >0 g). The dichotomization was done because the short-term intake variables showed pronounced right-skewed distributions. For the habitual intake analysis, the intake variables were kept as continuous variables, and regression analysis was applied.

The MUVR analysis had 2 initial steps: *1*) a test run was performed for each food group with PLS and RF with a low number of repetitions (*n*Rep = 6), and *2*) the modeling approach (PLS or RF) with the best performance, that is., highest Q^2 or lowest number of misclassifications, for the specific food group was selected. Both PLS and RF have been used in biomarker discovery in the past and have strengths and weaknesses [26,50]. However, this initial 2-step approach ensures that the best-performing modeling approach is selected for each specific food group. Furthermore, step 1 provided valuable information on whether PLS or RF could identify predictive metabolite features; if the initial model could not predict intake, that is, negative Q^2 or zero correctly classified consumers, no further analysis was performed for that food group.

For the main analysis, the parameter configuration was set to *n*Rep = 50 (number of repetitions), varRatio = 0.85 (ratio of variables retained per iteration), and *n*Outer = 8 (number of outer test segments) for all models, in accordance with the recommendations by Shi et al. [50]. We selected metabolite features from the minimal-optimal model, representing the minimal variable set necessary for optimal method performance. Furthermore, after the initial feature selection by RF or PLS, linear mixed models were applied for the final metabolomic feature selection. An unstructured covariance structure was assumed, and the subject ID, unique to each participant, was set as a random effect. In the linear mixed models, each preselected feature (dependent variable) was regressed on the dietary intake variable (independent variable) and adjusted for the covariates age, sex, country, batch, BMI *z*-score, and energy intake. These covariates were selected as a minimal necessary set of individual characteristics; Supplemental Figure 1 shows a directed acyclic graph showing the theoretical framework for the adjustment in the statistical models. To account for multiple hypothesis testing, the Benjamini–Hochberg procedure was applied to control the false discovery rate at 5%, considering results with a *q* value of <0.05 to be statistically significant. In the DONALD cohort, an identical statistical pipeline was applied (Supplemental Material).

Metabolite annotation

After the final feature selection, the list of metabolite features was sent back to IARC for annotation. The metabolite features were compared with the in-house database of analytical standards with a 10-ppm mass and a retention time tolerance of 0.25 min. in addition, the *m/z* values were searched in the human metabolome database [51] with a 10-ppm mass tolerance. The quality of the chromatographic peaks and spectra was inspected, and the plausibility of database candidates was assessed based on retention time, isotope pattern, adduct formation, and neutral losses. The best matching identities were confirmed by matching the MS/MS spectra and retention time from the metabolite and the corresponding standard. When standards were not available, MS/MS spectra were compared against those in mzCloud or METLIN [52]. The level of identification was determined as proposed by Sumner et al. [53].

Results

The baseline and follow-up characteristics were stratified by study sample, being presented in Table 1. At baseline, the median age for both study samples was 6.4 y, and almost half of the participants were female. Most of the participants in the short-term intake sample were from Germany, whereas most participants in the habitual intake sample were from Italy at baseline. In total, 11,397 metabolite features were analyzed in negative ionization mode and 16,559 metabolite features in positive ionization mode in the repeated urine samples. After the initial filtration (i.e., removal of features with >95% missing values), 1616 and 1567 metabolite features remained in negative ionization mode and 2055 and 1984 metabolite features in positive ionization mode, for the short-term and habitual intake samples, respectively (Supplemental Figures 2 and 3 show flow diagrams of the filtration process).

Identification of metabolites in the IDEFICS/I.Family cohort

Short-term intake

Table 2 presents an overview of the number of consumers for each food group and time point. After the initial MUVR test runs, the food groups potato crisps and jelly candy were excluded from the main MUVR analysis. Furthermore, for the food groups chocolate and nut spread and ice cream, the negative and positive ionization mode metabolite features were not included in the main MUVR analysis, respectively. For an overview of the test runs and main analysis with the MUVR algorithms, see Supplemental Tables 1 and 2. After MUVR and linear mixed model analyses, while accounting for repeated measures, 16 metabolite features were associated with chocolate candy intake and 8 metabolite features were associated with candy and sweet intake and were issued for annotation. Of the 24 metabolite features, 4 metabolites were annotated, namely theobromine,

TABLE 1
Baseline and follow-up characteristics of the study samples from the IDEFICS/I.Family cohort for short-term and habitual dietary intake.

	Habitual dietary intake study sample, $n = 599$			Short-term dietary intake study sample, $n = 296$		
	W1, $n = 597$	W2, $n = 596$	W3, $n = 595$	W1, $n = 116$	W2, $n = 105$	W3, $n = 223$
Median (range)						
Age, y	6.4 (2.1–9.3)	8.4 (4–11.1)	12.3 (8–15.2)	6.4 (2.7–9.3)	8.5 (4.0–11.1)	11.8 (8.2–15.2)
BMI z-score	0.42 (−2.79 to 5.07)	0.48 (−3.3 to 4.18)	0.66 (−2.13 to 3.64)	−0.01 (−1.80 to 2.34)	0.13 (−1.75 to 2.64)	0.48 (−2.13 to 3.64)
Energy intake (kcal)[1]	1642 (1136–2409)	1613 (900–2802)	1643 (779–2656)	1462 (213–2426)	1597 (266–3551)	1462 (274–4346)
n (%)						
Female	281 (47)	282 (47)	280 (47)	52 (45)	53 (51)	107 (48)
Country[2]						
Italy	288 (48)	287 (48)	288 (48)	3 (3)	7 (7)	46 (21)
Estonia	140 (23)	140 (23)	139 (23)	2 (2)	0	97 (43)
Belgium	12 (2)	12 (2)	12 (2)	10 (9)	3 (3)	6 (3)
Sweden	51 (9)	50 (8)	49 (8)	47 (41)	43 (41)	35 (16)
Germany	62 (10)	63 (11)	63 (11)	53 (46)	45 (43)	23 (10)
Hungary	30 (5)	30 (5)	30 (5)	1 (0.9)	5 (5)	5 (2)
Spain	14 (2)	14 (2)	14 (2)	0	2 (2)	11 (5)

Abbreviations: FFQ, food frequency questionnaire; IDEFICS, Identification and Prevention of Dietary- and Lifestyle-induced Health Effects in Children and Infants; NCI, United States National Cancer Institute; 24-HDR, 24-h dietary recall; 3d-WDR, 3-d weighted dietary record.
[1] Usual energy intake calculated from 24-HDR and FFQ data with NCI method; short-term energy intake from 24-HDR.
[2] Summed percentage over 100% due to rounding.

TABLE 2
Number of short-term consumers of the food groups in the IDEFICS/I.Family cohort by study time points.

Consumers	W1, $n = 116$	W2, $n = 105$	W3, $n = 223$
	n (%)		
Chocolate candy	34 (29)	27 (26)	57 (26)
Chocolate and nut spread	18 (16)	8 (8)	20 (9)
Potato crisps	4 (3)	6 (6)	17 (8)
Jelly candy	6 (5)	9 (9)	9 (4)
Candy and sweets	25 (22)	14 (13)	25 (11)
Cakes, puddings, and cookies	59 (51)	54 (51)	86 (39)
Ice cream	13 (11)	10 (10)	21 (9)

cyclo(L-prolyl-L-valyl), xanthosine, and octenoylcarnitine. Table 3 provides the overview of the annotated metabolites and Supplemental Tables 3 and 4 an overview of the selected metabolite features that could not be annotated.

The first sensitivity analysis showed an association between chocolate candy intake and theobromine and cyclo(L-prolyl-L-valyl), but not with xanthosine, and between candy and sweets intake and octenoylcarnitine. Additionally, some new metabolite features were identified for chocolate candy, jelly candy, chocolate and nut spread, candy and sweets, and cakes, puddings and cookies (Supplemental Tables 5 and 6). The second sensitivity analysis showed no attenuation of the associations between chocolate candy intake and theobromine, cyclo(L-prolyl-L-valyl), and xanthosine (results not shown).

Habitual intake

Table 4 depicts an overview of the median intake in grams per day for each food group and time point. After the initial MUVR test runs, the food group savory and fatty snacks was not selected for the main MUVR analysis. Results of the test runs and main analysis with the MUVR algorithms are provided in the Supplemental Tables 7 and 8. We accounted for repeated measures during MUVR and linear mixed model runs. The analyses showed that 45 metabolite features were associated with chocolate candy intake; 5 metabolite features were associated with candy and sweet intake; 1 metabolite feature was associated with crisp intake; 9 metabolite features were associated with the intake of cakes, puddings, and cookies; and 2 metabolite features were associated with ice cream intake. Of these metabolite features, 3 metabolites could be annotated. Same as for short-term intake, habitual intake of chocolate candy was associated with theobromine, cyclo(L-prolyl-L-valyl), and xanthosine. Additionally, intake of cakes, puddings, and cookies was associated with theobromine. Table 3 presents the overview of identified and annotated metabolites and Supplemental Tables 3 and 4 an overview of the selected metabolite features that could not be annotated.

Replication of metabolites in the DONALD cohort

During the initial MUVR analysis, only chocolate candy intake was moved forward to the final metabolite feature selection. Supplemental Table 9 presents an overview of the individual characteristics and the median intake for chocolate candy in the DONALD replication cohort.

In the independent DONALD cohort, chocolate candy intake was associated with 45 metabolite features. Of these metabolite features, 4 were annotated, namely theobromine, cyclo(L-prolyl-L-valyl), xanthosine, and 3-hydroxyphenylacetate (Table 5).

Discussion

Using data from a large European cohort of children and adolescents, we were able to identify potential biomarkers of short-term and habitual intake of sweet and fatty snacks. Importantly, 3 putative biomarkers of chocolate intake, namely theobromine, xanthosine, and cyclo(L-prolyl-L-valyl), were externally replicated in the German DONALD cohort.

Most studies in children have focused on identifying biomarkers of fruit and vegetable intake. Furthermore, many studies

TABLE 3
Overview of the identified short-term and/or habitual dietary intake-metabolite associations in the IDEFICS/I.Family cohort.

Food group	Intake type[1]	Coefficient[2]	SE	q[3]	Metabolite	Ionization mode	Mass	Identification level[4]	Regulation	HMDB ID	Metabolic pathway
Chocolate candy	Short-term	0.01	0.002	0.02	Theobromine	Positive	180.06473	1	Up	0002825	Caffeine metabolism
	Habitual	0.01	0.003	<0.001							
Cakes, puddings, and cookies	Habitual	0.01	0.002	0.01							
Chocolate candy	Short-term	0.01	0.002	0.01	Xanthosine	Positive	284.07565	1	Up	0000299	Purine metabolism
	Habitual	0.01	0.003	0.003		Negative	284.07584				
Chocolate candy	Short-term	0.01	0.002	0.03	Cyclo(L-prolyl-L-valyl)	Positive	196.12143	1	Up	0240493	NA
	Habitual	0.02	0.003	<0.001							
Candy and sweets	Short-term	-0.01	0.003	0.04	Octenoylcarnitine	Positive	285.19403	2	Down	NA	NA

[1] Short-term: dietary intake 1 or 2 d before urine collection; habitual: dietary intake calculated with the United States National Cancer Institute method.
[2] Coefficients are on the log and z scale and adjusted for age, sex, country, batch, BMI z-score, and energy intake.
[3] The q value is a p value that has been adjusted for the false discovery rate.
[4] Identification level, see Sumner et al. [53].

have used cross-sectional data or data from dietary interventions for biomarker identification [25]. Only 2 studies have used cohort data to identify BFIs in children [24,54]. Our study used longitudinal cohort data with the repeatedly measured urine metabolome to identify biomarkers of fatty and sweet snack intake in children. This approach highlights the novelty of this study and can contribute to the further use of cohort data in the field of biomarker identification.

Before discussing the results, it is important to highlight the main challenges of our methodologic approach. The analyses were based on large cohort studies but still depended on self-reported dietary intake data. In this study, the 24-HDR and FFQ were used to assess the dietary intake of the participants of the main cohort, methods often criticized for their questionable validity in accurately reflecting children's diets [4]. The 24-HDR, used as short-term intake measurement, likely introduces a bias toward the null, because sweet and fatty snack intakes might be underreported [4,55]. Nevertheless, efforts were made to validate the 24-HDR and FFQ, enhancing accurate dietary assessment [33,35,38]. Furthermore, the combination of 24-HDR and FFQ data in this study reduces biases and increases the accuracy of individual habitual dietary intake estimates [56,57]. Our approach, however, would not be suitable to capture biomarkers of rarely consumed foods, which would require feeding studies. However, this study focused on regularly consumed snacks. Only biomarkers with longer half-lives and foods frequently consumed would be useful for cohort studies that typically rely on single urine sample collections. We studied acute intake biomarkers as a proof of principle by linking 24-HDR with urine samples from the previous day. The successful replication of 3 biomarker candidates can also be seen as a proof of concept for our methodologic approach.

In this study, besides the 4 annotated metabolites, 62 unique metabolite features were selected but could not be annotated. Unfortunately, it is possible neither to provide any further information on the identity of these metabolite features nor to assess biological plausibility. This is common in untargeted metabolomics analysis due to the vast number of features detected by mass spectrometry [58]. Additionally, ~60 more unknowns were selected in the sensitivity analysis. The mass of these unknowns is provided in the Supplemental Material for future studies to use for reference.

The candy and sweets–related metabolite in the discovery cohort, octenoylcarnitine, belongs to the class of acylcarnitines and is a medium-chain acylcarnitine. Increased concentrations of octenoylcarnitine are associated with obesity and fatty acid metabolism disorders [59]. Only 2 studies have reported associations between diet and octenoylcarnitine [60,61]. In this study, we found a negative association between candy and sweet intake and octenoylcarnitine. According to the validation criteria by Dragsted et al. [62], a metabolite should increase in response to dietary intake, that is, there should be a positive association with dietary intake. Therefore, octenoylcarnitine is unlikely a relevant BFI of candy and sweet intake in our cohort.

Theobromine is an organic compound and belongs to the class of xanthines [63]. In humans, this metabolite is a product of caffeine breakdown by CYP1A2 in the liver [64]. Measured in blood, theobromine has a half-life of 6–8 h. Theobromine is found in the highest concentrations in cocoa products and in smaller concentrations in coffee and tea [63,65]. Accordingly, many studies have identified theobromine as a potential BFI of cocoa

TABLE 4
Median habitual dietary intake of the food groups in the IDEFICS/I.Family cohort by study time points.

Dietary intake (g/d)	W1, $n = 597$	W2, $n = 596$	W3, $n = 595$
	Median (range)		
Chocolate candy	7.8 (1.2–55.6)	6.8 (2.1–52.2)	11.4 (2.7–71.2)
Savory and fatty snacks	62.1 (30.3–119.4)	70.2 (2.5–153.8)	63.2 (21.2–153.7)
Candy and sweets	0.6 (0.2–27.1)	3.1 (0.5–21.4)	1.4 (0.1–39.8)
Crisps	1.3 (0.4–138.9)	1.2 (0.3–119.7)	6.2 (2.4–42.4)
Cakes, puddings, and cookies	65.3 (34.8–99.2)	60.8 (19.4–121.4)	55.0 (24.5–140.2)
Ice cream	4.1 (1.4–152.5)	16.0 (0.9–204.1)	11.9 (4.4–70.4)

TABLE 5
Overview of the annotated metabolites measured in negative and positive ionization mode associated with chocolate intake in the DONALD cohort—results from the replication analysis.

Food group	Coefficient[1]	SE	q[2]	Metabolite	Ionization mode	Mass	Identification level[3]	Regulation
Chocolate candy	0.02	0.003	<0.001	Theobromine	Positive	180.06467	1	Up
Chocolate candy	0.02	0.003	<0.001	Cyclo(L-prolyl-L-valyl)	Positive	196.12136	1	Up
Chocolate candy	0.01	0.003	<0.001	Xanthosine[4]	Positive	366.14252	1	Up
Chocolate candy	0.01	0.003	0.001	Xanthosine[4,5]	Positive	284.07553	1	Up
Chocolate candy	0.01	0.003	0.003	Xanthosine[4]	Positive	244.14232	1	Up
Chocolate candy	0.01	0.003	0.001	3-hydroxyphenylacetate	Negative	152.04745	2	Up
Chocolate candy	0.01	0.003	0.01	Xanthosine[4]	Negative	366.14261	1	Up
Chocolate candy	0.01	0.003	0.03	Xanthosine[4,5]	Negative	284.07574	1	Up

[1] Coefficients are on the log scale and adjusted for age, sex, BMI, and energy intake.
[2] The q value is a p value that has been adjusted for the false discovery rate.
[3] Identification level, see Sumner et al. [53].
[4] Metabolite features belong to the same metabolite.
[5] Main metabolite.

products [17]. However, Michielsen et al. [17] dismissed theobromine as a potential BFI of cocoa products, arguing that it is not specific to the consumption of cocoa products. Indeed, in our study, theobromine was also associated with intake of cakes, puddings, and cookies, but this may reflect the cocoa ingredients in these snacks.

Xanthosine is a purine nucleoside and an intermediate in the purine metabolism [66]. It is produced during the breakdown of theobromine, which is derived from caffeine [67]. Xanthosine is expected to be present in cocoa beans, and other foods but has not been quantified in these until now [66]. Several studies have associated methylxanthines, such as 7-methylxanthosine, which is the precursor of xanthosine, with cocoa (products) consumption [17]. However, our literature search found no evidence linking chocolate intake and xanthosine in previous literature.

Cyclo(L-prolyl-L-valyl) belongs to the class of α-amino acids and derivatives [68]. Few studies have been published about this metabolite. Nonetheless, it is linked to cocoa (products) and coffee intake [69]. One study found a correlation between cyclo(L-prolyl-L-valyl) and chocolate intake [70], whereas another study found a correlation between cyclo(L-prolyl-L-valyl) and habitual coffee intake [71]. Beyond these findings, we could not identify any study that reported an association between chocolate or cocoa (products) and cyclo(L-prolyl-L-valyl).

Theobromine, xanthosine, and cyclo(L-prolyl-L-valyl) may be potential candidate biomarkers of coffee and/or black tea intake, making them unsuitable for identifying chocolate intake. However, this issue is less relevant for younger children, who are typically nonconsumers of coffee and black tea, although it may be relevant for adolescents. A sensitivity analysis was performed by adding short-term coffee and tea intake as a covariate into the linear mixed model, which did not alter the results of the short-term chocolate intake analysis. Only very few children reported consuming coffee or tea (W1, $n = 16$; W2, $n = 16$; and W3, $n = 55$).

Notably, Europe is among the largest chocolate consumers globally, followed closely by the United States [72]. According to the European Food Safety Agency, chocolate (beverages) is the main source of caffeine intake for children aged 3–10 y [73]. Therefore, these BFIs could be valuable in assessing chocolate intake in children. From a public health perspective, the detection of metabolites from caffeine pathways in children consuming chocolate may be controversial. Future studies should evaluate the half-life and pharmacologic properties of xanthines like theobromine for children.

Indeed, the elimination half-life of the metabolites might be an important aspect. In the first sensitivity analysis, we restricted the short-term sample to participants with a 1-d interval between urine collection and dietary intake. We could identify theobromine, cyclo(L-prolyl-L-valyl), and octenoylcarnitine, as well as additional metabolites for 6 food groups not found in the main analysis. The closer timing between dietary intake and urine collection enabled us to detect associations between short-term dietary intake and metabolites that might have been cleared in those with a longer interval between intake and collection.

All 3 chocolate metabolites were replicated in the DONALD cohort, adding to the evidence that theobromine, xanthosine, and cyclo(L-prolyl-L-valyl) are potential biomarkers of chocolate intake in children and adolescents. In addition, 3-hydroxyphenylacetate, a metabolite not previously identified as a chocolate intake biomarker, was found in the replication cohort. The

differences between the main and replication cohort, such as dietary assessment methods (24-HDR compared with 3d-WDR), urine collection modes (morning urine compared with 24h-urine), and country of residence (European countries compared with Germany), could explain the different results. The 24-h urine collection may present higher validity and could capture a wider range of metabolites, whereas the timing of morning urine samples might miss some metabolites due to their elimination half-life [74]. Although 1 study suggested only small differences between the applied dietary assessment methods, the impact on the study outcome is still uncertain [75]. Furthermore, the country of residence may be an important factor and explains a part of the variance in the urine metabolome [24,76].

Lastly, we could not identify potential BFIs for all snack food groups we investigated. This may be due to the nonspecific nature of some of the food groups, for example, combining cakes, puddings, and cookies into 1 group; and the low number of consumers for certain items, like jelly candy. For an in-depth study of biomarkers of other snack foods, a more specific dietary assessment of the nature, ingredients, and brands of these convenience snack foods, along with higher sample sizes, would be recommended.

That said, the limitations of this study must be highlighted: The main limitation is the reliance on self-reported dietary intake, as discussed in the beginning. Additionally, the half-life of elimination may have affected the short-term analysis, where metabolite features were found for only 2 of 7 food groups. Although we sought to identify the best statistical pipeline for the available data, the use of machine learning (MUVR algorithms), and standard statistical techniques (linear mixed models) is not common practice in biomarker discovery [26,77]. Furthermore, the differential analysis mode for short-term intake (dichotomization and classification analysis) and habitual intake (continuous and regression analyses) could have influenced the number of metabolites related to short-term and habitual intake, although the number of features in both analyses was comparable.

This study has several strengths. We made use of 2 independent longitudinal cohorts of European children, which allowed us to evaluate the consistency of the results. The 3 chocolate metabolites were identified by a 2-step statistical analysis strategy, which was also applied in the DONALD cohort. The application of repeated double crossvalidation with unbiased variable selection algorithms provided a stable selection of metabolite features in the face of a large number of initial metabolite features [50]. Additionally, we applied linear mixed models to account for important individual characteristics. In both cohorts, the whole (food) metabolome was measured by untargeted metabolomics, covering a large number of available metabolites in the urine samples. Repeatedly measured food metabolome data were available in both cohorts. The laboratory and statistical analysis pipelines were streamlined for both cohorts, making the analysis approaches as identical as possible. Combining the FFQ and 24-HDR data with the NCI method potentially increased the precision of the predicted habitual dietary intakes over merely averaging dietary intakes from the repeated 24-HDR [40].

In conclusion, in this study, we analyzed untargeted metabolomics data measured in repeatedly collected urine samples. We were able to identify and replicate theobromine, xanthosine, and cyclo(L-prolyl-L-valyl) in 2 independent cohorts of children and adolescents. This approach is novel and demonstrates the potential use of cohort data for the identification of biomarkers. Nonetheless, further research is needed to assess the validity and specificity of these potential chocolate BFIs in children, especially in light of coffee and tea intake. Controlled feeding studies with targeted metabolomics measurements would be needed for further validation.

Acknowledgments

We specially thank Dr. Augustin Scalbert who initiated the project collaboration as former head of the biomarkers group at IARC. This work was done as part of the IDEFICS (http://www.idefics.eu) and I.Family studies (http://www.ifamilystudy.eu/) and the DONALD study. We gratefully acknowledge the financial support of the European Commission within the Sixth RTD Framework Programme Contract No. 016181 (FOOD), and the Seventh RTD Framework Programme Contract No. 266044. The DONALD study is financially supported by the Ministry of Science and Research of North Rhine-Westphalia, Germany.

Author contributions

The authors' responsibilities were as follows – AF, UN, IH, PK-R, JR: designed research; SDH, MH, IH, LL, DM, LAM, PR, TV, UN: conducted research; JR, MM, DA, PK-R: performed laboratory analysis, metabolomics data preprocessing, and metabolite annotation; JG: performed the statistical analysis in the discovery cohort; LY: made significant contributions toward the statistical analysis in the discovery cohort; RF: provided guidance on the statistical analysis in the discovery cohort; SM: performed statistical analysis in the replication cohort; KO: revised and supervised statistical analysis in the replication cohort; JG: wrote paper; AK: supervised and made significant contributions to the writing process of the paper; JG, AF: had primary responsibility for final content; and all authors: have read and approved the final manuscript.

Conflict of interest

AF reports financial support was provided by German Research Foundation. PK-R reports financial support was provided by French National Research Agency. All other authors report no conflicts of interest.

Funding

This study was funded by the German Research Foundation (DFG project number 406710821) and the Agence Nationale de la Recherche (ANR project number ANR-18-CE92-0060). This work was done as part of the IDEFICS (http://www.idefics.eu) and I.Family studies (http://www.ifamilystudy.eu/) and the DONALD study. We gratefully acknowledge the financial support of the European Commission within the Sixth RTD Framework Programme Contract No. 016181 (FOOD), and the Seventh RTD Framework Programme Contract No. 266044. The DONALD study is financially supported by the Ministry of Science and Research of North Rhine-Westphalia, Germany.

Data availability

Data described in the manuscript, code book, and analytic code will be made available upon request pending application and approval.

Disclaimer

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer /World Health Organization.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tjnut.2024.09.026.

References

[1] P. Maruvada, J.W. Lampe, D.S. Wishart, D. Barupal, D.N. Chester, D. Dodd, et al., Perspective: dietary biomarkers of intake and exposure-exploration with omics approaches, Adv. Nutr. 11 (2) (2020) 200–215, https://doi.org/10.1093/advances/nmz075.
[2] A. de la Hunty, J. Buttriss, J. Draper, H. Roche, G. Levey, A. Florescu , et al., UK Nutrition Research Partnership (NRP) workshop: forum on advancing dietary intake assessment, Nutr. Bull. 46 (2) (2021) 228–237, https://doi.org/10.1111/nbu.12501.
[3] A. Afshin, P.J. Sur, K.A. Fay, L. Cornaby, G. Ferrara, J.S. Salama, Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017, Lancet 393 (10184) (2019) 1958–1972, https://doi.org/10.1016/S0140-6736(19)30041-8.
[4] M.B. Livingstone, P.J. Robson, J.M. Wallace, Issues in dietary intake assessment of children and adolescents, Br. J. Nutr. 92 (Suppl 2) (2004) S213–S222, https://doi.org/10.1079/bjn20041169.
[5] S.A. Bingham, Urine nitrogen as a biomarker for the validation of dietary protein intake, J. Nutr. 133 (3) (2003) 921S–924S.
[6] T.S. Collese, A.C.F. De Moraes, T. Rendo-Urteaga, L.A. Luzia, P.H.d.C. Rondó, D.M.L. Marchioni, et al., The validity of children's fruit and vegetable intake using plasma vitamins A, C, and E: the SAYCARE study, Nutrients 11 (8) (2019) 1815.
[7] M. Jenab, N. Slimani, M. Bictash, P. Ferrari, S.A. Bingham, Biomarkers in nutritional epidemiology: applications, needs and new horizons, Hum. Genetics. 125 (5) (2009) 507–525, https://doi.org/10.1007/s00439-009-0662-5.
[8] A. Lovell, R. Bulloch, C.R. Wall, C.C. Grant, Quality of food-frequency questionnaire validation studies in the dietary assessment of children aged 12 to 36 months: a systematic literature review, J. Nutr. Sci. 6 (2017) e16, https://doi.org/10.1017/jns.2017.12.
[9] Q. Gao, G. Praticò, A. Scalbert, G. Vergères, M. Kolehmainen, C. Manach, et al., A scheme for a flexible classification of dietary and health biomarkers, Genes Nutr 12 (2017) 34, https://doi.org/10.1186/s12263-017-0587-x.
[10] 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. 99 (6) (2014) 1286–1308, https://doi.org/10.3945/ajcn.113.076133.
[11] E. Shibutami, T. Takebayashi, A scoping review of the application of metabolomics in nutrition research: the literature survey 2000–2019, Nutrients 13 (11) (2021) 3760.
[12] Q. Sun, K.A. Bertrand, A.A. Franke, B. Rosner, G.C. Curhan, W.C. Willett, Reproducibility of urinary biomarkers in multiple 24-h urine samples, Am. J. Clin. Nutr. 105 (1) (2017) 159–168, https://doi.org/10.3945/ajcn.116.139758.
[13] T. Rafiq, S.M. Azab, K.K. Teo, L. Thabane, S.S. Anand, K.M. Morrison , et al., Nutritional metabolomics and the classification of dietary biomarker candidates: a critical review, Adv. Nutr. 12 (6) (2021) 2333–2357.
[14] E.M. Brouwer-Brolsma, B. Brandl, M.E.C. Buso, T. Skurk, C. Manach, Food intake biomarkers for green leafy vegetables, bulb vegetables, and stem vegetables: a review, Genes Nutr 15 (1) (2020) 7, https://doi.org/10.1186/s12263-020-00667-z.
[15] C. Cuparencu, G. Pratico, L.Y. Hemeryck, P. Harsha, S. Noerman, C. Rombouts, et al., Biomarkers of meat and seafood intake: an extensive literature review, Genes Nutr 14 (1) (2019) 30, https://doi.org/10.1186/s12263-019-0656-4.
[16] M. Garcia-Aloy, P.J.M. Hulshof, S. Estruel-Amades, M.C.J. Oste, M. Lankinen, J.M. Geleijnse, et al., Biomarkers of food intake for nuts and vegetable oils: an extensive literature search, Genes Nutr 14 (2019) 21, https://doi.org/10.1186/s12263-019-0628-8.
[17] C. Michielsen, E. Almanza-Aguilera, E.M. Brouwer-Brolsma, M. Urpi-Sarda, L.A. Afman, Biomarkers of food intake for cocoa and liquorice (products): a systematic review, Genes Nutr 13 (2018) 13, https://doi.org/10.1186/s12263-018-0610-x.
[18] L.H. Munger, M. Garcia-Aloy, R. Vazquez-Fresno, D. Gille, A.R.R. Rosana, A. Passerini, et al., Biomarker of food intake for assessing the consumption of dairy and egg products, Genes Nutr 13 (2018) 18, https://doi.org/10.1186/s12263-018-0615-5.
[19] G. Pratico, Q. Gao, C. Manach, L.O. Dragsted, Biomarkers of food intake for Allium vegetables, Genes Nutr 13 (2018) 12, https://doi.org/10.1186/s12263-018-0624-4.
[20] M. Ulaszewska, N. Vázquez-Manjarrez, M. Garcia-Aloy, R. Llorach, F. Mattivi, L.O. Dragsted, et al., Food intake biomarkers for apple, pear, and stone fruit, Genes Nutr 13 (2018) 29, https://doi.org/10.1186/s12263-018-0620-8.
[21] R. Vazquez-Fresno, A.R.R. Rosana, T. Sajed, T. Onookome-Okome, N.A. Wishart, D.S. Wishart, Herbs and spices—biomarkers of intake based on human intervention studies—a systematic review, Genes Nutr 14 (2019) 18, https://doi.org/10.1186/s12263-019-0636-8.
[22] N. Vázquez-Manjarrez, M. Ulaszewska, M. Garcia-Aloy, F. Mattivi, G. Praticò, L.O. Dragsted, et al., Biomarkers of intake for tropical fruits, Genes Nutr 15 (1) (2020) 11, https://doi.org/10.1186/s12263-020-00670-4.
[23] X.M. Zhou, Q. Gao, G. Pratico, J. Chen, L.O. Dragsted, Biomarkers of tuber intake, Genes Nutr 9 (2019) 15, https://doi.org/10.1186/s12263-019-0631-0.
[24] C.E. Lau, A.P. Siskos, L. Maitre, O. Robinson, T.J. Athersuch, E.J. Want, et al., Determinants of the urinary and serum metabolome in children from six European populations, BMC Med 16 (1) (2018) 202, https://doi.org/10.1186/s12916-018-1190-8.
[25] L. Yuan, S. Muli, I. Huybrechts, U. Nöthlings, W. Ahrens, A. Scalbert , et al., Assessment of fruit and vegetables intake with biomarkers in children and adolescents and their level of validation: a systematic review, Metabolites 12 (2) (2022) 126.
[26] M.C. Playdon, A.D. Joshi, F.K. Tabung, S. Cheng, M. Henglin, A. Kim , et al., Metabolomics analytics workflow for epidemiological research: perspectives from the Consortium of Metabolomics Studies (COMETS), Metabolites 9 (7) (2019) 145, https://doi.org/10.3390/metabo9070145.
[27] E.D. Clarke, M.E. Rollo, K. Pezdirc, C.E. Collins, R.L. Haslam, Urinary biomarkers of dietary intake: a review, Nutr. Rev. 78 (5) (2020) 364–381, https://doi.org/10.1093/nutrit/nuz048.
[28] W. Ahrens, K. Bammann, A. Siani, K. Buchecker, S. De Henauw, L. Iacoviello, et al., The IDEFICS cohort: design, characteristics and participation in the baseline survey, Int. J. Obes. 35 (1) (2011) S3–S15.
[29] W. Ahrens, A. Siani, R. Adan, S. De Henauw, G. Eiben, W. Gwozdz , et al., Cohort profile: the transition from childhood to adolescence in European children-how I.Family extends the IDEFICS cohort, Int. J. Epidemiol. 46 (5) (2017) 1394–1395j, https://doi.org/10.1093/ije/dyw317.
[30] A. Hebestreit, T. Intemann, A. Siani, S. De Henauw, G. Eiben, Y.A. Kourides, et al., Dietary patterns of European children and their parents in association with family food environment: results from the I. family study, Nutrients 9 (2) (2017) 126.
[31] W. Ahrens, K. Bammann, I. Pigeot, The IDEFICS/I.Family studies: design and methods of a large European child cohort, in: K. Bammann, L. Lissner, I. Pigeot, W. Ahrens (Eds.), Instruments for health surveys in children and adolescents, Springer, 2019, pp. 1–24.
[32] T.J. Cole, T. Lobstein, Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity, Pediatr. Obes. 7 (4) (2012) 284–294.
[33] A. Hebestreit, M. Wolters, H. Jilani, G. Eiben, V. Pala, Web-based 24-h dietary recall: the SACANA program, in: K. Bammann, L. Lissner, I. Pigeot, W. Ahrens (Eds.), Instruments for health surveys in children and adolescents, Springer, 2019, pp. 77–102.
[34] V. Pala, L.A. Reisch, L. Lissner, Dietary behaviour in children, adolescents and families: the eating habits questionnaire (EHQ), in: K. Bammann, L. Lissner, I. Pigeot, W. Ahrens (Eds.), Instruments for health surveys in children and adolescents, Springer, 2019, pp. 103–133.
[35] T. Intemann, I. Pigeot, S. De Henauw, G. Eiben, L. Lissner, V. Krogh , et al., Urinary sucrose and fructose to validate self-reported sugar

intake in children and adolescents: results from the I.Family study, Eur. J. Nutr. 58 (3) (2019) 1247–1258.
[36] A. Hebestreit, G. Barba, S. De Henauw, G. Eiben, C. Hadjigeorgiou, É. Kovács, et al., Cross-sectional and longitudinal associations between energy intake and BMI z-score in European children, Int. J. Behav. Nutr. Phys. Act. 13 (1) (2016) 23.
[37] I. Iglesia, T. Intemann, P. De Miguel-Etayo, V. Pala, A. Hebestreit, M. Wolters, et al., Dairy Consumption at snack meal occasions and the overall quality of diet during childhood. Prospective and cross-sectional analyses from the IDEFICS/I.Family cohort, Nutrients 12 (3) (2020) 642.
[38] I. Huybrechts, C. Börnhorst, V. Pala, L. Moreno, G. Barba, L. Lissner , et al., Evaluation of the Children's Eating Habits Questionnaire used in the IDEFICS study by relating urinary calcium and potassium to milk consumption frequencies among European children, Int. J. Obes. 35 (1) (2011) S69–S78.
[39] A.F. Subar, K.W. Dodd, P.M. Guenther, V. Kipnis, D. Midthune, M. McDowell, et al., The food propensity questionnaire: concept, development, and validation for use as a covariate in a model to estimate usual food intake, J. Am. Diet. Assoc. 106 (10) (2006) 1556–1563, https://doi.org/10.1016/j.jada.2006.07.002.
[40] V. Kipnis, D. Midthune, D.W. Buckman, K.W. Dodd, P.M. Guenther, S.M. Krebs-Smith, et al., Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes, Biometrics 65 (4) (2009) 1003–1010.
[41] J. Peplies, K. Günther, A. Gottlieb, A. Lübke, K. Bammann, W. Ahrens, Biological samples—standard operating procedures for collection, shipment, storage and documentation, in: K. Bammann, L. Lissner, I. Pigeot, W. Ahrens (Eds.), Instruments for health surveys in children and adolescents, Springer, 2019, pp. 57–76.
[42] J. Peplies, K. Günther, K. Bammann, A. Fraterman, P. Russo, T. Veidebaum, et al., Influence of sample collection and preanalytical sample processing on the analyses of biological markers in the European multicentre study IDEFICS, Int. J. Obes (Lond). 35 (Suppl 1) (2011) S104–S112, https://doi.org/10.1038/ijo.2011.41.
[43] A. Buyken, U. Alexy, M. Kersting, T. Remer, [The DONALD cohort. An updated overview on 25 years of research based on the Dortmund Nutritional and Anthropometric Longitudinally Designed study], Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 55 (6–7) (2012) 875–884.
[44] I. Perrar, U. Alexy, U. Nöthlings, Cohort profile update–overview of over 35 years of research in the Dortmund Nutritional and Anthropometric Longitudinally Designed (DONALD) study, Eur. J. Nutr. 63 (2023) 1–14.
[45] K.T. Do, S. Wahl, J. Raffler, S. Molnos, M. Laimighofer, J. Adamski , et al., Characterization of missing values in untargeted MS-based metabolomics data and evaluation of missing data handling strategies, Metabolomics 14 (10) (2018) 128, https://doi.org/10.1007/s11306-018-1420-2.
[46] R Core Team, R:, a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2022.
[47] H. Wickham, M. Averick, J. Bryan, W. Chang, L.D.A. McGowan, R. François, et al., Welcome to the Tidyverse, J. Open Source Softw. 4 (43) (2019) 1686.
[48] J. Pinheiro, D. Bates, R Core Team, nlme: linear and nonlinear mixed effects models, R package version 3.1-157, R Foundation for Statistical Computing, Vienna, Austria, 2022.
[49] C. Brunius, MUVR: multivariate methods with Unbiased Variable selection in R, R package version 0.0.975, R Foundation for Statistical Computing, Vienna, Austria, 2022.
[50] L. Shi, J.A. Westerhuis, J. Rosén, R. Landberg, C. Brunius, Variable selection and validation in multivariate modelling, Bioinformatics 35 (6) (2019) 972–980.
[51] D.S. Wishart, A. Guo, E. Oler, F. Wang, A. Anjum, H. Peters, et al., HMDB 5.0: the Human Metabolome Database for 2022, Nucleic Acids Res 50 (D1) (2022) D622–D631, https://doi.org/10.1093/nar/gkab1062.
[52] C.A. Smith, G.O. Maille, E.J. Want, C. Qin, S.A. Trauger, T.R. Brandon, et al., METLIN: a metabolite mass spectral database, Ther. Drug Monit. 27 (6) (2005) 747–751.
[53] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, et al., Proposed minimum reporting standards for chemical analysis, Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI), Metabolomics 3 (3) (2007) 211–221, https://doi.org/10.1007/s11306-007-0082-2.
[54] M. Prasad, H.-M. Takkinen, L. Uusitalo, H. Tapanainen, M.-L. Ovaskainen, G. Alfthan, et al., Carotenoid intake and serum concentration in young Finnish children and their relation with fruit and vegetable consumption, Nutrients 10 (10) (2018) 1533.
[55] L. Garden, H. Clark, S. Whybrow, R.J. Stubbs, Is misreporting of dietary intake by weighed food records or 24-hour recalls food specific? Eur. J. Clin. Nutr. 72 (7) (2018) 1026–1034, https://doi.org/10.1038/s41430-018-0199-6.
[56] M. Looman, H.C. Boshuizen, E.J.M. Feskens, A. Geelen, Using enhanced regression calibration to combine dietary intake estimates from 24 h recall and FFQ reduces bias in diet–disease associations, Public Health Nutr 22 (15) (2019) 2738–2746, https://doi.org/10.1017/S1368980019001563.
[57] L.S. Freedman, D. Midthune, L. Arab, R.L. Prentice, A.F. Subar, W. Willett, et al., Combining a food frequency questionnaire with 24-hour recalls to increase the precision of estimation of usual dietary intakes—evidence from the validation studies pooling project, Am. J. Epidemiol. 187 (10) (2018) 2227–2232, https://doi.org/10.1093/aje/kwy126.
[58] Z. Zhou, M. Luo, H. Zhang, Y. Yin, Y. Cai, Z.J. Zhu, Metabolite annotation from knowns to unknowns through knowledge-guided multi-layer metabolic networking, Nat. Commun. 13 (1) (2022) 6656, https://doi.org/10.1038/s41467-022-34537-6.
[59] M. Dambrova, M. Makrecka-Kuka, J. Kuka, R. Vilskersts, D. Nordberg, M.M. Attwood, et al., Acylcarnitines: nomenclature, biomarkers, therapeutic potential, drug targets, and clinical trials, Pharmacol. Rev. 74 (3) (2022) 506–551, https://doi.org/10.1124/pharmrev.121.000408.
[60] N.V. Khodorova, A. Rietman, D.N. Rutledge, J. Schwarz, J. Piedcoq, S. Pilard, et al., Urinary medium-chained acyl-carnitines sign high caloric intake whereas short-chained acyl-carnitines sign high-protein diet within a high-fat, hypercaloric diet in a randomized crossover design dietary trial, Nutrients 13 (4) (2021) 1191.
[61] G. Gürdeniz, M. Uusitupa, K. Hermansen, M.J. Savolainen, U. Schwab, M. Kolehmainen, et al., Analysis of the SYSDIET Healthy Nordic Diet randomized trial based on metabolic profiling reveal beneficial effects on glucose metabolism and blood lipids, Clin. Nutr. 41 (2) (2022) 441–451, https://doi.org/10.1016/j.clnu.2021.12.031.
[62] L.O. Dragsted, Q. Gao, A. Scalbert, G. Vergères, M. Kolehmainen, C. Manach, et al., Validation of biomarkers of food intake-critical assessment of candidate biomarkers, Genes Nutr 13 (2018) 14, https://doi.org/10.1186/s12263-018-0603-9.
[63] Human Metabolome Database, Theobromine [Internet], 2023. Available from: https://hmdb.ca/metabolites/HMDB0002825#references.
[64] S. Martínez-López, B. Sarriá, M. Gómez-Juaristi, L. Goya, R. Mateos, L. Bravo-Clemente, Theobromine, caffeine, and theophylline metabolites in human plasma and urine after consumption of soluble cocoa products with different methylxanthine contents, Food Res. Int. 63 (2014) 446–455, https://doi.org/10.1016/j.foodres.2014.03.009.
[65] IARC Working Group on the Evaluation of Carcinogenic Risks to Humans Coffee, tea, mate, methylxanthines and methylglyoxal, IARC Working Group on the evaluation of carcinogenic risks to humans: Theorbomine, IARC, Lyon, France, 1991.
[66] Human Metabolome Database, Xanthosine [Internet], 2023. Available from: https://hmdb.ca/metabolites/HMDB0000299.
[67] F. Sato, 2.26—plant alkaloid engineering, in: H.-W. Liu, TP Begley (Eds.), Comprehensive natural products III, Elsevier, Oxford, 2020, pp. 700–755.
[68] Human Metabolome Database, cyclo(L-Prolyl-L-valyl) [Internet], 2023. Available from: https://foodb.ca/compounds/FDB093683.
[69] H. Sook Chung, S.Y. Lee, Modification of ginseng flavors by bitter compounds found in chocolate and coffee, J. Food Sci. 77 (6) (2012) S202–S210, https://doi.org/10.1111/j.1750-3841.2012.02716.x.
[70] D.Y. Low, S. Lefèvre-Arbogast, R. González-Domínguez, M. Urpi-Sarda, P. Micheau, M. Petera, et al., Diet-related metabolites associated with cognitive decline revealed by untargeted metabolomics in a prospective cohort, Mol. Nutr. Food Res. 63 (18) (2019) e1900177, https://doi.org/10.1002/mnfr.201900177.
[71] J.A. Rothwell, P. Keski-Rahkonen, N. Robinot, N. Assi, C. Casagrande, M. Jenab, et al., A metabolomic study of biomarkers of habitual coffee intake in four European countries, Mol. Nutr. Food Res. 63 (22) (2019) 1900659, https://doi.org/10.1002/mnfr.201900659.
[72] H.C. Alberts, J. Cidell, Chocolate consumption, manufacturing, and quality in Europe and North America, in: M.P. Squicciarini, J. Swinnen (Eds.), The economics of chocolate, Oxford University Press, Oxford, United Kingdom, 2015, pp. 119–133.
[73] EFSA Panel on Dietetic Products, Nutrition, and Allergies, Scientific opinion on the safety of caffeine, EFSA J 13 (5) (2015) 4102, https://doi.org/10.2903/j.efsa.2015.4102.
[74] J. Goerdten, L. Yuan, I. Huybrechts, V. Neveu, U. Nöthlings, W. Ahrens, et al., Reproducibility of the blood and urine exposome: a systematic

literature review and meta-analysis, Cancer Epidemiol. Biomarkers Prev. 31 (9) (2022) 1683–1692.
[75] H. Xue, M. Yang, Y. Liu, R. Duan, G. Cheng, X. Zhang, Relative validity of a 2-day 24-hour dietary recall compared with a 2-day weighed dietary record among adults in South China, Nutr. Diet 74 (3) (2017) 298–307, https://doi.org/10.1111/1747-0080.12315.
[76] A. Trimigno, B. Khakimov, F. Savorani, L. Tenori, V. Hendrixson, A. Čivilis, et al., Investigation of variations in the human urine metabolome amongst European populations: an exploratory search for biomarkers of people at risk-of-poverty, Mol. Nutr. Food Res. 63 (1) (2019) 1800216, https://doi.org/10.1002/mnfr.201800216.
[77] S. Noerman, J.K. Virtanen, M. Lehtonen, C. Brunius, K. Hanhineva, Serum metabolites associated with wholegrain consumption using nontargeted metabolic profiling: a discovery and reproducibility study, Eur. J. Nutr. 62 (2023) 713–726.

3.4 Publication 4

Association of ultra-processed foods intake with untargeted metabolomics profiles in adolescents and young adults in the DONALD cohort study.
https://doi.org/10.1016/j.tjnut.2024.09.023

The Journal of Nutrition 154 (2024) 3255–3265

American Society for Nutrition
Excellence in Nutrition Research and Practice

JN THE JOURNAL OF NUTRITION

journal homepage: https://jn.nutrition.org/

Genomics, Proteomics, and Metabolomics

Association of Ultraprocessed Foods Intake with Untargeted Metabolomics Profiles in Adolescents and Young Adults in the DONALD Cohort Study

Samuel Muli [1], Annika Blumenthal [1], Christina-Alexandra Conzen [1], Maike Elena Benz [1], Ute Alexy [1], Matthias Schmid [2], Pekka Keski-Rahkonen [3], Anna Floegel [4], Ute Nöthlings [1,*]

[1] *Unit of Nutritional Epidemiology, Department of Nutrition and Food Sciences, University of Bonn, Bonn, Germany;* [2] *Institute for Medical Biometry, Informatics and Epidemiology (IMBIE), University Hospital Bonn, Bonn, Germany;* [3] *International Agency for Research on Cancer (IARC), Lyon, France;* [4] *Section of Dietetics, Faculty of Agriculture and Food Sciences, Hochschule Neubrandenburg, Neubrandenburg, Germany*

A B S T R A C T

Background: High consumption of ultraprocessed foods (UPFs) continues to draw significant public health interest because of the associated negative health outcomes. Metabolomics can contribute to the understanding of the biological mechanisms through which UPFs may influence health.

Objectives: To investigate urine and plasma metabolomic biomarkers of UPF intake in adolescents and young adults.

Methods: We used data from the Dortmund Nutritional and Anthropometric Longitudinally Designed study to investigate cross-sectional associations of UPF intake with concentrations of urine metabolites in adolescents using 3d weighed dietary records (3d-WDR) and 24-h urine samples (n = 339), and associations of repeatedly assessed UPF intake with concentrations of circulating plasma metabolites in young adults with 3–6 3d-WDRs within 5 y preceding blood measurement (n = 195). Urine and plasma samples were analyzed using mass spectrometry-based metabolomics. Biosample-specific metabolite patterns (MPs) were determined using robust sparse principal components analysis. Multivariable linear regression models were applied to assess the associations of UPF consumption (as a percentage of total food intake in g/d) with concentrations of individual metabolites and MP scores.

Results: The median proportion of UPF intake was 22.0% [interquartile range (IQR): 12.3, 32.9] in adolescents and 23.2% (IQR: 16.0, 31.6) in young adults. We identified 42 and 6 UPF intake-associated metabolites in urine and plasma samples, respectively. One urinary MP, "xenobiotics and amino acids" [β = 0.042, 95% confidence interval (CI): 0.014, 0.070] and 1 plasma MP, "lipids, xenobiotics, and amino acids" (β = 0.074, 95% CI: 0.031, 0.117) showed positive association with UPF intake. Both patterns shared 29 metabolites, mostly of xenobiotic metabolism.

Conclusions: We identified urine and plasma metabolites associated with UPF intake in adolescents and young adults, which may represent some of the biological mechanisms through which UPFs may influence metabolism and health.

Keywords: ultraprocessed foods, dietary biomarkers, metabolomics, metabolites, metabolite patterns

Introduction

Industrial food processing, which combines ingredients and additives, is important for improving food safety, nutritional access, and in reducing food waste across the supply chain [1]. Ultraprocessed foods (UPFs) as defined by the NOVA classification system [2], includes a broad range of industrially processed food products such as soft drinks, flavored yogurts, packaged snacks, confectionery, pasta and pizza dishes, processed meat products, instant noodles, soups and sauces, among other ready-to-heat or -eat food products [2–4]. Broadly, most UPFs contain a mix of food-derived or reconstituted ingredients and other industrially isolated components such as lactose, casein, gluten, whey, hydrogenated oils, variety of sugars (e.g., high fructose corn syrup), cosmetic additives such as colorants, dyes, flavors enhancers, bulking agents, emulsifiers among others,

Abbreviations: 3d-WDR, 3-d weighed dietary record; CI, confidence interval; DONALD, Dortmund Nutritional and Anthropometric Longitudinally Designed; MP, metabolite pattern; PCA, principal component analysis; UMF, unprocessed or minimally processed food; UPF, ultraprocessed food.

* Corresponding author. *E-mail address:* noethlings@uni-bonn.de (U. Nöthlings).

https://doi.org/10.1016/j.tjnut.2024.09.023
Received 2 August 2024; Received in revised form 16 September 2024; Accepted 22 September 2024; Available online 25 September 2024

mostly of exclusive industrial use [2,3,5]. These formulations provide convenience, affordability, and enhance shelf-life and sensory properties [2].

The consumption of UPFs has been rising, contributing to more than half of the total daily energy intake in some countries, for example, the United States [6]. In 22 European countries, UPF intake varies markedly, ranging from 14% to 44%, as reported in a recent study [7]. An increasing number of epidemiologic studies suggest that high consumption of UPFs is associated with increased health risks—including obesity [8], cardiometabolic diseases [4,9], chronic kidney disease [10], cancer [9,11], irritable bowel syndrome [12], and depression [13]. Some of these health risks have been shown to be particularly significant for higher intake of animal-based products and sweetened beverages [9].

Several mechanisms have been proposed to explain ways through which UPFs may influence health. These include poor nutrient profiles: excessive added sugars, salts, unhealthy fats, and high energy density but low in protein and dietary fiber [2,14]. Furthermore, the physical and chemical properties associated with industrial processing, ingredients, and their by-products may also contribute to these increased health risks [4]. Nevertheless, the debate, widespread disagreement, and uncertainty on the links between UPFs and increased health risks still exists [15–19].

Untargeted metabolomics is a promising approach for investigating the relationship between the consumption of UPFs and health status, as dietary intakes elicit metabolic changes that can be related to health indicators. Urine and blood matrices may reflect different aspects of dietary intake and metabolic changes, with urine more reliably reflecting short-term changes in concentrations of diet-responsive metabolites, and blood a more stable overview of an individual's metabolic state [20]. These insights could improve our understanding of pathways through which UPF may affect health, encouraging more nuanced discourses on these biological mechanisms. Indeed, some metabolomics-based studies have investigated the links between individual foods within the UPFs such as sweetened beverages [21], processed red meats [22,23], and metabolic health. Such studies are valuable, particularly in identifying biomarkers associated with the intake of specific foods. However, considering UPFs as an aggregate dietary pattern may more accurately reflect actual dietary habits, as foods and nutrients are typically consumed in combinations [24–26].

So far, the associations of various dietary patterns and metabolomics profiles have been extensively described [27], but studies on UPF-metabolome associations are markedly fewer. This gap is evident across all age groups, with only a handful of studies conducted in adults [10,28,29] and in younger populations [30,31]. Therefore, our study investigates the association between UPF intake and untargeted urine and plasma metabolomics profiles in adolescents and young adults from a well-characterized German cohort.

Methods

Study design and population

The Dortmund Nutritional and Anthropometric Longitudinally Designed (DONALD) study is an open dynamic cohort on children and young individuals residing in Dortmund, Germany. Since 1985, 30–40 healthy infants are recruited annually during their first year and are regularly assessed and followed up until adulthood. Regular assessments include dietary intake, anthropometric measurements, urine sample collection (from the ages 3–4 y), blood samples (from 18 y), medical, lifestyle, and other sociodemographic data [32].

The DONALD study was approved by the Ethics Committee of the University of Bonn (ethics numbers: 098/06 and 185/20). All procedures and assessments followed ethical standards of the Declaration of Helsinki, with written informed consent from parents and from adolescents from the age of 16 y.

Study sample

The current analyses included 2 analytic samples termed adolescent urine $n = 339$, and young adult plasma $n = 195$, previously described in [33]. In brief, the adolescent urine sample included individuals who provided a 3-d weighed dietary record (3d-WDR) and a single 24-h urine sample (age at urine sample, 14.9–18.4 y). The young adult plasma sample included individuals who completed 3 or more 3d-WDRs within the 5-y period preceding a single blood measurement (age at blood sample, 18.0–21.9 y). The 2 analytic samples had an overlap of 139 participants. A study flowchart is provided in Supplemental Figure 1.

Dietary assessment

In the 3d-WDRs, the study participants, or with parental assistance, weighed all foods and beverages consumed and leftovers to the nearest 1 g using electronic food scales. Semi-quantitative estimates (e.g., portion sizes, cups, or spoons) were acceptable if weighing the food was not possible such as for meals consumed away from home. Information on recorded food products, including brands, ingredients, declared nutrients, and methods of preparation were also collected. Each product recorded for the first time receives its own entry in an internally maintained and regularly updated food composition in-house database (LEBTAB) [34]. The nutritional profiles of staple foods were derived from German food composition tables, whereas the caloric and nutrient content of packaged food products (e.g., processed items, convenience meals, and snacks) were determined through recipe simulation from their ingredients and nutrients labels.

Food and beverage groupings

All food and beverage items recorded by the participants were assigned into 1 of the NOVA categories according to the purpose, nature, and extent of their processing [2]: NOVA-1 [unprocessed or minimally processed foods (UMFs)], NOVA-2 (processed culinary ingredients), NOVA-3 (processed foods), and NOVA-4 (UPF). Food items with unclear NOVA classification were documented and classified on the basis of internal consensus and the Federal Ministry of Food and Agriculture guideline for spices and other seasonings [35]. A summary of food and beverage groups according to the NOVA system is provided in Supplemental Table 1.

This study primarily focused on the UPFs. Daily intakes were calculated as individual means from the 3d-WDR. We defined the long-term UPF intake variable as the mean intake across all 3d-WDRs within the 5-y period preceding blood draw. In our

main analysis, we calculated the proportion of UPF intake as a percentage of total weight of food and beverages consumed in grams. This weight-based ratio instead of the energy-based ratio acknowledges food and beverages that provide low or zero calories [11,36,37] as well as nonnutritive ingredients and additives that may be used in food processing [37].

Other covariate assessment

Experienced nurses conducted anthropometric measurements following standardized procedures. BMI (kg/m^2) was calculated from these measurements. Leisure time physical activity was assessed using a questionnaire adapted from the validated Adolescent Physical Activity Recall Questionnaire [38]. Participants estimated the time spent on a range of organized and unorganized sports over the past 12 mo, and the reported activities quantified in metabolic equivalent of task-hours per week. Self-reported alcohol use and smoking status were assessed using a questionnaire, and participants categorized as current, former, and never for each of these lifestyle factors. The covariate data represent measurements closest to or on the date of biosample collection.

Urine and blood samples

Following a standardized protocol, participants collected their 24-h urine samples on the third day of the 3-d dietary assessment. These were then stored in sterile, preservative-free plastic containers at temperatures below −12°C. Upon transfer to the DONALD study center, the urine samples were stored at −22°C until processed. Fasting blood sample were drawn, centrifuged at 4°C for 15 min (3100 U/min, 2000 × *g*), aliquoted and stored at −80°C. EDTA plasma was used in this analysis. Detailed procedures for urine and blood samples are provided [32].

Metabolite profiling

Metabolon Inc. conducted untargeted metabolomics analysis on the urine and plasma samples using ultrahigh performance liquid chromatography-tandem mass spectroscopy. Metabolon followed their standardized protocol for sample handling, raw data extraction, and peak identification and analysis as outlined in their procedures [39]. For plasma samples, Metabolon applied both metabolomics and lipidomics approaches. Overall, 1407 metabolites were annotated in urine and 1042 features in plasma samples. We provide a detailed description of the metabolomics procedures for both urine and blood samples in the Supplemental Methods.

Statistical analyses

Descriptive statistics

Characteristics of study participants were expressed as medians (25th and 75th percentile) for continuous variables and frequencies (percentages) for categorical variables.

Processing of metabolite data

We performed mechanism aware imputation of missing values in 2 steps. First, we applied a novel method that combines particle swarm optimization (to search for metabolite concentration thresholds and the proportion of low concentration deletions) and extreme gradient boosting as classifier method for mechanism underlying each missing value, following procedures provided by Yuan et al. [40]. These were implemented in Python using NumPy, Pandas, scikit-learn, and XGBoost libraries. Our data predominantly showed missing not at random values in urine (86.1%) and in plasma (83.2%) samples. Therefore, metabolites with >20% missing data were excluded according to the "80% rule" [41], and the rest were imputed by quantile regression imputation of left-censored data using MetImp 1.2 [42]. Batch normalization were conducted using *ber* bagging method implemented in the dbnorm R package [43]. Subsequently, the data were natural log-transformed, mean centered, and scaled to unit variance.

Deriving metabolite patterns

We used the robust sparse principal component analysis (Robust SPCA) to compute the metabolite patterns (MPs) because of better interpretability of its components through sparse vectors, as the loadings are determined from a subset of the original variables, and its robustness to outlying observations [44]. We implemented these steps using the sparsepca R package. The optimal PCA components retained were determined by scree plots using the PCAtools R package.

Associations of UPF with metabolites and MPs

Using multivariable linear regression, we regressed *1*) each of the single metabolites and *2*) each of the MPs on the UPF intake, adjusting for age, sex, energy intake, BMI, physical activity, smoking, and alcohol status. The plasma models were additionally adjusted for the time difference between dietary assessments and blood draw (time difference = age at blood draw − mean age of dietary assessments) and the number of dietary assessments per participant. We applied the Benjamini–Hochberg procedure to control the false discovery rate at 5% within each set of regression analyses. For the main results, we assessed model assumptions (i.e., normality of residuals, linearity, and homogeneity of variance) using the performance R package.

Missing covariates

Considering the DONALD's longitudinal design, we first applied backward filling for "never" alcohol intake (or smokers) to fill missing data for earlier time points. We then imputed the rest of missing data for physical activity, alcohol use, and smoking status (Table 1) using the K-nearest neighbor algorithm, with 10 nearest neighbors on the basis of other nonmissing covariate data. These were implemented using the VIM R package.

Additional analyses

Considering differences in UPF variable definition in literature, such as absolute intakes, weight-based ratio, and energy-based ratio [9–11,28–31,36,37], we conducted secondary analysis to compare our main results (weight-based ratio) with absolute (g/d) and energy-based ratio (energy from UPFs as a percentage of total energy intake). We also computed correlation (Pearson) between UPF and UMF intakes to investigate the hypothesis that a higher UPF consumption is related to reduced UMF intake [2,4]. Lastly, we performed sensitivity analyses on urine samples, $n = 260$, and plasma, $n = 137$, after excluding potentially implausible 3d-WDR reporting on the basis of sex- and age-specific thresholds for underreporting [45].

All statistical analyses were conducted using Python (v3.8) and R (v4.1.3).

TABLE 1
Basic characteristics of the study participants.

	n	Adolescent urine ($N = 339$)	*n*	Young adult plasma ($N = 195$)
Sex: female	339	166 (49.0)	195	108 (55.4)
Age at biosample collection (y)	339	18.0 (17.0, 18.1)	195	18.1 (18.1, 18.2)
BMI (kg/m^2)	339	21.9 (19.9, 24.0)	195	22.2 (20.1, 24.5)
UPF, % total food intake (g/d)	339	22.0 (12.3, 32.9)	195	23.2 (16.0, 31.6)
UPF, % TEI	339	42.0 (32.2, 52.0)	195	45.0 (36.8, 50.8)
Energy intake (TEI, Kcal/d)	339	2126.9 (1748.5, 2582.1)	195	1978.1 (1697.0, 2390.1)
3d-WDR assessments	339	1.0	195	4.0 (4.0, 5.0)
Physical activity (MET-h /w)	215	34.0 (14.1, 54.8)	184	30.1 (12.1, 52.9)
Smoking status	211		142	
Never		155 (73.5)		98 (69.0)
Former		23 (10.9)		21 (14.8)
Current		33 (15.6)		23 (16.2)
Alcohol status	179		153	
Never		24 (13.4)		20 (13.1)
Former		27 (15.1)		31 (20.3)
Current		128 (71.5)		102 (66.7)

Abbreviations: 3d-WDR, 3-d weighed dietary records, MET-h /w, metabolic equivalent of task-hours per week; TEI, total energy intake; UPF, ultraprocessed food.
Data are presented as *n* (%) and median (25th and 75th percentile) for categorical and continuous variables, respectively. Differences in *n* are because of missing data values.
The young adult label reflects age at blood sample collection (min-max. 18.0–21.9 y) but the repeated dietary assessments were conducted over the 5 y preceding the blood draw.

Results

Descriptive characteristics

The adolescent urine samples (49.0% female) and young adult plasma samples (55.4% female) had median ages of 18.0 and 18.1 y, with median BMIs of 21.9 and 22.2, respectively. The median proportion of UPF intake to total food intake by weight was 22.0% (IQR: 12.3%, 32.9%) in adolescents and 23.2% (IQR: 16.0%, 31.6%) in young adults (Table 1). The foods with highest mean contribution to total UPF intake in both analytic samples were sweetened beverages (nondairy soft drinks) and convenient, ready-to-heat or -eat food products, contributing 31.3% and 11.5% (adolescents) and 28.1% and 11.2% (young adults), respectively (Figure 1). Regarding energy intake, however, sweets, chocolates, and ice cream; cereals and industrial breads; and processed meats and sausages had the highest energy contribution to the total calorie intake from the UPFs in both adolescents and young adults (Supplemental Figure 2).

Associations between UPF intake and urine metabolites

In adolescent urine samples, of the 42 urine metabolites identified in our fully adjusted model, 21 metabolites were positively associated with UPF intake (Table 2). In accordance with the Dragsted et al. [46,47] framework for evaluating food biomarker plausibility, which suggests that biomarkers of intake should demonstrate dose–response relationship (i.e., increase in human sample with higher food intake), in this section, we highlight the positive associations of UPF intake and urine metabolites and provide all associations observed in Supplemental Table 2.

Among the known metabolites, higher UPF intake was associated with higher concentrations of indoxyl glucuronide $\beta = 0.013$ [95% confidence intervals (CIs): 0.007, 0.019] and other partially characterized glucuronides: glucuronide of C10H18O2 (1), $\beta = 0.015$ (0.008, 0.021); glucuronide of C10H14O2 (2), $\beta = 0.015$ (0.008, 0.021); glucuronide of C10H18O2 (7), $\beta = 0.014$

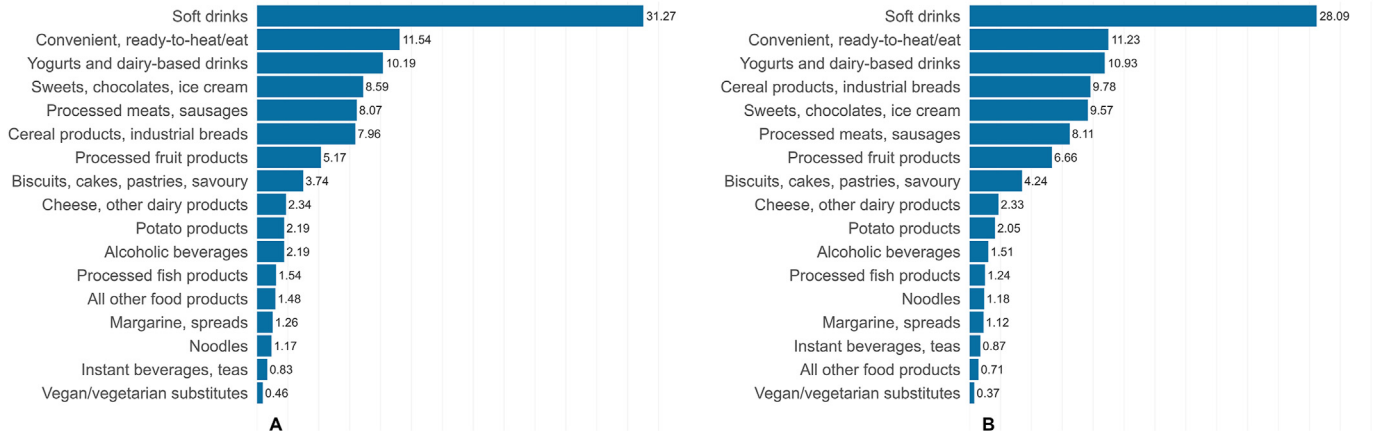

FIGURE 1. Mean contribution of various food groups to the total UPF consumption, as a percentage of the total weight (g/d) in (A), adolescent urine and (B), young adults' plasma analytic samples. UPF, ultraprocessed food.

TABLE 2
Regression estimates of the associations between UPF intake and urine metabolites in adolescents ($n = 339$).

Metabolite	Model 1, β (95% CI)	Model 2, β (95% CI)	Model 3, β (95% CI)
X-17679	0.026 (0.021, 0.032)[1]	0.026 (0.020, 0.032)[1]	0.027 (0.021, 0.033)[1]
X-19497	0.013 (0.007, 0.019)[1]	0.014 (0.008, 0.020)[1]	0.016 (0.010, 0.022)[1]
Glucuronide of C10H18O2 (1)[2]	0.014 (0.007, 0.020)[1]	0.014 (0.007, 0.020)[1]	0.015 (0.008, 0.021)[1]
Glucuronide of C10H14O2 (2)[2]	0.013 (0.007, 0.019)[1]	0.013 (0.007, 0.020)[1]	0.015 (0.008, 0.021)[1]
X-11478	0.012 (0.006, 0.018)[1]	0.013 (0.006, 0.019)[1]	0.014 (0.007, 0.020)[1]
Glucuronide of C10H18O2 (7)[2]	0.012 (0.006, 0.018)[1]	0.012 (0.006, 0.019)[1]	0.014 (0.007, 0.021)[1]
Indoxyl glucuronide	0.005 (–0.001, 0.011)	0.010 (0.004, 0.016)[1]	0.013 (0.007, 0.019)[1]
Glucuronide of C10H18O2 (8)[2]	0.013 (0.007, 0.019)[1]	0.012 (0.006, 0.019)[1]	0.014 (0.007, 0.020)[1]
N,N-dimethylalanine	0.010 (0.004, 0.016)[1]	0.012 (0.006, 0.018)[1]	0.013 (0.007, 0.019)[1]
3-Hydroxy-N6,N6,N6-trimethyl-L-lysine[2]	0.012 (0.006, 0.018)[1]	0.012 (0.006, 0.018)[1]	0.012 (0.006, 0.018)[1]
1-methylhistamine	0.011 (0.005, 0.016)[1]	0.011 (0.005, 0.017)[1]	0.012 (0.006, 0.018)[1]
Glucuronide of C10H18O2 (9)[2]	0.011 (0.005, 0.017)[1]	0.011 (0.005, 0.017)[1]	0.012 (0.006, 0.019)[1]
X-25442	0.011 (0.005, 0.017)[1]	0.013 (0.007, 0.019)[1]	0.012 (0.006, 0.019)[1]
X-17825	0.009 (0.003, 0.015)[1]	0.011 (0.004, 0.017)[1]	0.011 (0.005, 0.017)[1]
X-24345	0.010 (0.004, 0.016)[1]	0.012 (0.006, 0.018)[1]	0.012 (0.005, 0.018)[1]
X-17358	0.011 (0.005, 0.017)[1]	0.010 (0.004, 0.017)[1]	0.011 (0.005, 0.018)[1]
1,6-anhydroglucose	0.010 (0.004, 0.016)[1]	0.011 (0.005, 0.017)[1]	0.011 (0.004, 0.018)[1]
3-indoxyl sulfate	0.003 (–0.003, 0.009)	0.008 (0.002, 0.014)	0.010 (0.004, 0.016)[1]
Glycerophosphorylcholine (GPC)	0.005 (–0.001, 0.011)	0.009 (0.002, 0.015)	0.011 (0.004, 0.017)[1]
Glycolate (hydroxyacetate)	0.012 (0.006, 0.018)[1]	0.010 (0.004, 0.016)[1]	0.010 (0.004, 0.016)[1]
6-Bromotryptophan	0.011 (0.005, 0.017)[1]	0.010 (0.004, 0.016)[1]	0.010 (0.004, 0.016)[1]

Abbreviations: β, regression estimate; CI, confidence interval; UPF, ultraprocessed food.
Model 1: Unadjusted.
Model 2: Adjusted for age, sex, BMI, and energy intake.
Model 3: Adjustments in model 2 and physical activity, alcohol and smoking status.
Each model was run independently (i.e., separately not based on statistical significance in previous model), overall significance was based on model 3.
[1] Statistically significant results (false discovery rate-adjusted q value < 0.05). Only positive associations are summarized; full results are given in Supplemental Table 2.
[2] Indicates a compound that has not been confirmed based on authentic chemical standard, but Metabolon are confident in its identity. The structural identities of "X-" followed by a number (e.g., X-17679) are unknown.

(0.007, 0.021), glucuronide of C10H18O2 (8), $\beta = 0.014$ (0.007, 0.020). Other metabolites were N, N-dimethylalanine, $\beta = 0.013$ (0.007, 0.019); 1-methylhistamine, $\beta = 0.012$ (0.006, 0.018); 3-Hydroxy-N6,N6,N6-trimethyl-L-lysine, $\beta = 0.012$ (0.006, 0.018); glycerophosphorylcholine, $\beta = 0.011$ (0.004, 0.017); 1,6-anhydroglucose, $\beta = 0.011$ (0.004, 0.018); 3-indoxyl sulfate, $\beta = 0.010$ (0.004, 0.016); glycolate, $\beta = 0.010$ (0.004, 0.016); and 6-bromotryptophan, $\beta = 0.010$ (0.004, 0.016) (Table 2).

There were also structurally unknown metabolites whose concentrations positively correlated with UPF intake, namely X-17679, $\beta = 0.027$ (0.021, 0.033); X-19497, $\beta = 0.016$ (0.010, 0.022); X-11478, $\beta = 0.014$ (0.07, 0.020), among others (Table 2).

Associations between UPF intake and plasma metabolites

In young adult plasma samples, UPF intake was associated with 6 metabolites in our fully adjusted model after corrections for multiple testing (Table 3). Of these, UPF intake was associated with elevated concentrations of 4-hydroxyglutamate,

TABLE 3
Regression estimates of the associations between UPF intake and plasma metabolites in young adults ($N = 195$).

Metabolite	Model 1, β (95% CI)	Model 2, β (95% CI)	Model 3, β (95% CI)
Homostachydrine[1]	–0.021 (–0.031, –0.011)[2]	–0.025 (–0.035, –0.015)[2]	–0.024 (–0.034, –0.013)[2]
4-Hydroxyglutamate	0.017 (0.007, 0.028)	0.023 (0.013, 0.033)[2]	0.021 (0.011, 0.031)[2]
3-CMPFP	–0.019 (–0.029, -0.009)	–0.020 (–0.031, –0.010)[2]	–0.020 (–0.031, –0.009)[2]
X-11372	0.020 (0.009, 0.030)	0.018 (0.008, 0.028)[2]	0.019 (0.009, 0.030)[2]
X-23639	–0.015 (–0.025, –0.004)	–0.020 (–0.030, –0.009)[2]	–0.020 (–0.031, –0.010)[2]
X-24951	0.018 (0.008, 0.028)	0.019 (0.009, 0.029)[2]	0.019 (0.009, 0.030)[2]

Abbreviations: β, regression estimate; 3-CMPFP, 3-carboxy-4-methyl-5-pentyl-2-furanpropionate; CI, confidence interval; UPF, ultraprocessed food.
Model 1: Unadjusted.
Model 2: Adjusted for age, sex, BMI, energy intake, number of dietary assessments, and time difference between dietary assessment and blood draw.
Model 3: Adjustments in model 2 and physical activity, alcohol, and smoking status.
[1] Indicates a compound that has not been confirmed based on authentic chemical standard, but Metabolon are confident in its identity. The structural identities of "X-" followed by a number (e.g., X-11372) are unknown.
[2] Statistically significant results (false discovery rate-adjusted q value <0.05).

$\beta = 0.021$ (0.011, 0.031) and 2 structurally unknown metabolites, X-11372, $\beta = 0.019$ (0.009, 0.030) and X-24951, $\beta = 0.019$ (0.009, 0.030).

Associations between UPF intake and urine and plasma MPs

In adolescent urine samples, 25 MPs, explaining 61.7% of urine metabolite variation, were analyzed in relation to UPF intake (Supplemental Table 3). Four MPs (MP7, MP9, MP10, and MP18) were associated with UPF intake in the minimally adjusted model (Table 4). Two of these associations were independent of lifestyle factors and multiple testing correction: MP9, $\beta = 0.042$ (0.014, 0.070) and MP7, $\beta = -0.063$ (–0.092, –0.034). In brief, the MP9, consisting of $n = 214$ metabolites, was dominated by metabolites in the xenobiotic super pathway (63 metabolite of subclasses: food components, xanthine metabolism, chemicals and drugs, and benzoate metabolism), 64 structurally unknown metabolites, 40 amino acids, 16 lipids, 11 partially characterized molecules (particularly, glucuronides of C8H14O2, C8H14O2, C8H16O2, C10H18O2, C12H22O4, C12H22O3, and C14H26O4), and the rest were spread across nucleotides, cofactors and vitamins, energy, and peptides. Therefore, on the basis of known metabolic pathways, we named this urinary pattern "xenobiotics and amino acids" MP. The MP7, consisting of $n = 281$ metabolites, was dominated by unknown metabolites ($n = 74$), amino acids ($n = 70$), lipids ($n = 51$), and xenobiotics ($n = 41$) among other metabolite classes. We named this urinary pattern, "amino acids, lipids, and xenobiotics" MP.

We compared the 2 urinary MPs 9 and 7 and found 46 common metabolites, mostly showing opposite direction of PCA loadings, which also possibly reflects the results observed in multivariable linear regression models (i.e., UPF's positive association with MP9 and inverse association with MP7). Of these common metabolites, 17 represented xenobiotic metabolism (subclasses: food component, xanthine metabolism, and chemicals) and 13 amino acids (primarily involved in glycine, serine, and threonine; tryptophan; and alanine and aspartate metabolism) and other unknown metabolites and pathways. Extended tables of metabolites represented in MP9 and MP7 are provided in Supplemental Tables 4 and 5, with the top metabolites contributing to the variation and their loadings summarized in Supplemental Figures 3 and 4, respectively.

In young adult plasma samples, 19 MPs explaining 55.0% of the plasma metabolite variation were analyzed with the UPF intake (Supplemental Table 6). Four MPs (MP1, MP6, MP8, and MP17) were associated with UPF intake in the minimally adjusted model (Table 4). Of these, only MP8 was associated with UPF intake in our fully adjusted model after corrections for multiple testing, $\beta = 0.074$ (0.031, 0.117). This MP8 had $n = 216$ metabolites, dominated by lipids ($n = 86$), unknown metabolites ($n = 45$), xenobiotics ($n = 39$) amino acids ($n = 28$), and the rest spread across cofactors and vitamins, peptides, and partially characterized molecules. Similarly, on the basis of the known biochemical pathways, we named this pattern "lipids, xenobiotics, and amino acids" MP. An extended table of this MP is provided in Supplemental Table 7 and its top metabolites and weights in Supplemental Figure 5.

Lastly, we compared the similarity of the UPF-positively associated MPs across biological matrices (urine and plasma, i.e., urinary MP9 reflecting short-term intake and plasma MP8, reflecting repeated, long-term intake). We found 29 common metabolites, mostly xenobiotics (n =15) of subclasses food components, benzoate, xanthine, and drug/chemical pathways, and the rest were mostly amino acids and lipids.

Secondary and sensitivity analyses

In adolescent urine samples, most metabolites generally showed comparable results for both absolute and energy-based UPF variables, with few exceptions. For example, saccharin, a common noncaloric ingredient in sweetened beverages, was not statistically significant even before correcting for multiple statistical tests for the energy-based UPF, $\beta = 0.005$ (–0.003, 0.012) but was significant in absolute UPF intake, $\beta = 0.0002$ (0.0000, 0.0004) (Supplemental Table 8). The UPF-associated urinary MP9 was statistically significant with absolute and energy-based UPF, whereas the MP7 was statistically significant with energy-based UPF but not with the absolute UPF model (Supplemental Table 9).

In young adult plasma samples, most associations observed with the weight-based ratio were also replicated using absolute UPF intake and energy-based UPF (Supplemental Table 10). The

TABLE 4
Regression estimates of associations between UPF intake and urine and plasma metabolite patterns.

Analytic sample	MP	Model 1, β (95% CI)	Model 2, β (95% CI)	Model 3, β (95% CI)
Adolescent urine	MP7	–0.062 (–0.090, –0.034)[1]	–0.064 (–0.092, –0.036)[1]	–0.063 (–0.092, –0.034)[1]
	MP9	0.046 (0.20, 0.072)[1]	0.046 (0.019, 0.073)[1]	0.042 (0.014, 0.070)[1]
	MP10	0.022 (–0.003, 0.047)	0.036 (0.011, 0.061)[1]	0.029 (0.003, 0.055)
	MP18	–0.020 (–0.042, 0.002)	–0.032 (–0.055, –0.010)[1]	–0.032 (–0.056, –0.008)
Young adult plasma				
	MP1	–0.021 (–0.100, 0.058)	–0.081 (–0.134, –0.029)[1]	–0.064(–0.119, –0.010)
	MP6	0.047 (0.005, 0.088)	0.063 (0.023, 0.104)[1]	0.049 (0.008, 0.091)
	MP8	0.065 (0.026, 0.105)[1]	0.072 (0.031, 0.113)[1]	0.074 (0.031, 0.117)[1]
	MP17	–0.040 (–0.071, –0.010)	–0.052 (–0.083, –0.021)[1]	–0.038 (–0.070, –0.007)

Abbreviations: β, regression estimate; CI, confidence interval; MP, metabolite pattern; UPF, ultraprocessed food.
Model 1: Unadjusted.
Model 2: Adolescent urine – adjusted for age, sex, BMI, and energy intake.
Model 2: Young adult plasma – adjusted for age, sex, BMI, energy intake, number of dietary assessments, and time difference between dietary assessment and blood draw.
Model 3: Adjustments in model 2 and physical activity, alcohol and smoking status.
The MPs were analyzed and labeled separately for urine and plasma samples as MP1 to MPn.
[1] Statistical significance (false discovery rate-adjusted q value <0.05). Full results are given in Supplemental Tables 3 and 6.

UPF-related plasma MP8 was statistically significant with all UPF variable specifications (Supplemental Table 11).

Regarding the correlation between UPF and UMF intakes, we found a strong negative correlation of these intakes in both adolescent urine, $r = -0.88$ (–0.90, –0.86), and young adult plasma, $r = -0.95$ (–0.96, –0.95), samples.

In our sensitivity analyses on potentially implausible intakes, excluding possibly underreported intakes showed comparable results to those obtained from the main analysis using the entire analytic samples regardless of the UPF variable specification, urinary patterns (Supplemental Table 12) and plasma patterns (Supplemental Table 13).

Discussion

Using untargeted metabolomics, we investigated cross-sectional associations of UPF intake with urine metabolites and associations of repeatedly assessed UPF intake with plasma metabolites in a cohort of free-living adolescents and young adults. Our results suggest that the intake of UPF is reflected in the urine and plasma metabolome, through diverse biochemical pathways such as xenobiotics, amino acids, and lipids pathways, including alterations of microbiome-derived and other endogenous metabolites.

Noteworthy single-metabolite associations in urine included indoxyl glucuronide and several other partially characterized glucuronides. Glucuronidation is a major pathway for detoxification and elimination of exogenous substances, predominantly drugs, chemicals, dietary substances, and endogenous compounds such as hormones [48]. A recent controlled feeding trial also reported UPF-related changes in indoxyl glucuronide concentrations and various glucuronides of C10H18O2 in 24-h urine [28]. Similarly, we previously reported these associations (glucuronide of C10H14O2, glucuronide of C10H18O2) with sweetened beverages in 24-h urine samples [33], the food group with highest contribution to the UPF intake in our study. The specific mechanisms for dietary-related glucuronidation are not entirely defined, but the gut microbiome plays a key role in modulating microbial transformation of dietary substrates and glucuronide levels [48] and biosynthesis of microbial metabolites [49].

Indeed, 3-indoxyl sulfate, a protein-bound uremic toxin, and 6-bromotryptophan, both positively associated with UPF intake are microbiome-derived tryptophan metabolites. This association of UPF intake with indoxyl sulfate was also observed in adults [28]. These metabolites may be important because of the suggested roles of the UPFs and western-style diets in inducing gut microbial dysbiosis [50–52], and the associations of microbiome-related metabolites with health outcomes. In other studies, higher indoxyl sulfate concentrations were positively linked with various health problems: psychic anxiety [53], cognitive impairment [54], and neuroinflammation and oxidative stress [55].

Our finding of a positive association of UPFs with 1-methylhistamine is also consistent with a previous study on gut metabolites and microbial compositions, which observed that children on a western-style diet had elevated concentrations of 1-methylhistamine compared with those on a Mediterranean-style diet [56]. The mechanisms for elevated levels of 1-methylhistamine with UPF intake are unclear and could reflect histamine release as an allergic response [57], food histamine intolerance [58], or even microbiological contamination of certain foods [59].

The N,N-dimethylalanine, positively associated with the UPF in this study and similarly with sweetened beverages in our previous study [33], may be important in various biological processes. For instance, the urinary N,N-dimethylalanine was inversely associated with 3 measures of adiposity in adolescents [33], and lower concentrations of plasma N,N-dimethylalanine were observed in adults with type 2 diabetes in another study [60]. There is, however, limited literature on this metabolite.

Consistent with the study by O'Connor et al. [28], we observed that higher UPF intakes were associated with lower concentrations of metabolites known to reflect minimally processed or certain whole foods. Our single-metabolite models showed that higher UPF intake was linked to lower levels of hydroquinone sulfate, a marker of pear intake [61]; dopamine 3-O-sulfate, a marker for banana intake [62,63]; 2-acetamidophenol sulfate (HPAA sulfate), linked to whole grains [64]; 3,5-dihydroxybenzoic acid, also associated with intake of whole-grain cereals [47,65]; allantoin, a purine derivative found in cow milk [66]; caffeic acid sulfate, a polyphenol abundant in coffee beans [62]; ferulic acid 4-sulfate, a metabolite of ferulic and caffeic acid found in fruits, whole-grain cereals, and coffee [67] and 3-methyladipate, a metabolite of phytanic acid breakdown, found in meat, dairy fat, and fatty fish [68]. Taken together, these data suggest lower intake of these foods in diets rich in UPF.

The single-metabolite associations with UPFs were also captured in the urinary "xenobiotics and amino acids" pattern, MP9; whose composition includes nutrient- and nonnutrient-related metabolites. Many of these metabolites were reported in a previous feeding study [28]. Metabolites of xanthine metabolism, for example, caffeine and theobromine were also related to UPF intake in another study [10], whereas concentrations of theobromine, 7-methylxanthine, and 3-methylxanthine were elevated with UPF intake [28]. In our previous study including the same participants [33], we observed that these caffeine-related metabolites more likely reflected caffeine in the sweetened beverages independent of other dietary sources of caffeine. Theobromine is a well-known component of cocoa and present in chocolate and other cocoa containing foods. Thus, considering the main contributing foods to the UPF intake in our study, these metabolites and their variation (i.e., direction of their PCA loadings) relative to those known to reflect whole foods, make this pattern compellingly relevant to the UPF.

Furthermore, one of the distinctive characteristics of UPFs, according to the NOVA system, is the use of industrial food additives such as flavoring and preservative agents. In MP9, metabolites reflecting these include vanillate (4-hydroxy-3-methoxybenzoate), a widely used vanilla food flavor [69] and naringenin 7 glucuronide, metabolite of naringenin, an industrial flavoring agent extracted from grapefruit [70]. We, however, note that other food sources of naringenin exist, such as citrus fruits and fruit juices [71]. In addition, various forms of benzoates (e.g., sodium benzoate) are commonly used as preservatives in packaged foods [72]. Metabolites of benzoate metabolism pathway in the UPF-related MP9 were 3-methyl catechol sulfate, 4-ethylcatechol sulfate, 2-ethylphenylsulfate, and o-cresol.

In relation to long-term UPF intake and plasma metabolites, we observed far fewer associations compared with urine using the single-metabolite modeling approach. This might suggest generally weak individual plasma metabolite associations with UPF intake, with only a few remaining significant after correction for multiple testing. These differences could also reflect the biological characteristics of the sample matrices. Urine is a more reliable matrix for investigating short-term response to dietary intakes and detoxification pathways, whereas blood matrix more reliably reflects endogenous metabolism of food from the gut to the liver and blood [20]. An interesting result in our single-metabolite models was the association of higher UPF intake with elevated concentration levels of 4-hydroxyglutamate (4-hydroxy-L-glutamic acid). This association was recently observed in spot urine, 24-h urine, and plasma samples in a controlled feeding study by O'Connor et al. [28]. Given that 4-hydroxyglutamate is implicated in metabolic syndrome [73] and pre-eclampsia [74,75], its mechanistic link with the UPFs could be an interesting research target regarding diet–health relationships.

The UPF-plasma MP8 largely reflected lipid metabolites. Differences and variation of dietary modulation of urine and blood metabolome are anticipated [20], and some differences in our results may also reflect some of the long-term changes associated with UPF intake. Importantly, the common metabolites across the urine and plasma MPs positively linked to UPF were mostly involved in xenobiotic metabolism of food-related components, xanthine metabolism, benzoates, sulfites, and metabolites of potential contaminants or exposures to chemicals or drugs. Therefore, although unsupervised machine learning methods such as the PCA are exploratory and multiple biological interpretations may exist for observed exposure associations, collectively, the common biochemical pathways and their joint combinations as captured in these urine and plasma MPs reinforce their potential relevance to UPF intake.

In summary, our study provides further evidence and insights into the growing concern of the potential health effects of UPFs. Our data showed a strong negative correlation between intakes of UPF and UMF, suggesting that UPFs displace whole foods that should be the basis of the diet according to the NOVA system. These results are consistent with established literature and long-held perspectives on public health nutrition [2,4]. In line, our study extends this evidence to molecular level, as shown in our explorative single-metabolite models and MPs. We also note that the UPFs were associated with changes in gut microbiota metabolites (e.g., indoxyl sulfate and 6-bromotryptophan). Mechanisms for gut bacteria–xenobiotics interplay are well described [76], and western-style diets [51,52,77] or food additives themselves [50] are suggested to induce gut microbiome dysbiosis. Thus, our results suggest multiple ways through which regular intakes of UPFs may influence health—from nutritional displacement to alterations of the gut microbiota composition, and possibly other unintended but undetermined effects of industrial food additives and formulations. Some of these metabolites, which seem metabolically or physiologically important, could be putative links between the consumption of UPFs and health status.

This study has several strengths. The DONALD study design allows for regular, repeated dietary assessments on the same individuals that have enabled assessment of long-term UPF intake. The DONALD's 3d-WDRs, compared with other methods such as food-frequency questionnaires, collects dietary data at the food item level rather than at food group level, enabling more accurate assigning of their NOVA groups. Using urine and blood—the most popular biological matrices—and their complementary nature enabled us to investigate short-term and long-term metabolomics profiles of UPF intake. This approach has the potential to provide more insights on potentially transient as well as sustained metabolomics alterations related to dietary intakes. Besides, by combining single-metabolite models with the MP approaches, we identified individual metabolites as well as broader MPs that may mechanistically reflect the UPF-associated perturbations of the metabolome. We opted for non-targeted metabolomics approaches given that targeting single or multiple selected pathways might not optimally reflect the complex UPF influence on the metabolome.

We acknowledge several limitations in this study. Self-reported dietary assessments are subject to random and systematic errors. We performed sensitivity analysis on potential implausible dietary reporting, although these checks are based on energy intakes and only identifies potential underreporting or over-reporting. Another challenge relates to the uncertainty in classification of some multi-ingredient foods into processed (NOVA-3) or the UPFs (NOVA-4). Despite reaching consensus for food groupings based on multiple guidelines, some discrepancies may still exist. Furthermore, some of the metabolites consistently associated with UPF intakes in single-metabolite models and MPs were of unknown structural identities, limiting our understanding of their biological functions. Lastly, the DONALD cohort consists of a largely homogeneous, urban population with a higher socioeconomic status than the general German population [32]. This homogeneity could influence their food choices and dietary habits. Nonetheless, given that this study replicated findings of other studies from different populations and regions, suggests reasonable generalizability of our findings.

In conclusion, we identified individual metabolites and MPs that reflect UPF intake in urine and plasma samples of adolescents and young adults. These findings add to the growing literature on complementary assessment of UPF intake and reports some of the underlying biological mechanisms through which these foods may affect metabolism and health. Besides, the extensive processing of food resulting in UPFs has generated considerable clinical and public health interest and become mired in controversy. Our data suggest that the complex and heterogeneous nature of UPFs may be gleaned from the metabolome.

Acknowledgments

We thank all DONALD study participants and their families for participating in this study and the staff at the DONALD study center for data collection.

Author contributions

The authors' responsibilities were as follows – UN, AF, SM: conceptualization and research design; AB, CAC, MEB, UA, SM: compiled dietary data; PK-R: conducted research; SM: analyzed data and wrote the first draft; MS: statistical supervision; SM,

UN: had primary responsibility for final content, and all authors: read and approved the final manuscript.

Conflict of interest

The authors report no conflicts of interest. Where authors are identified as personnel of the International Agency for Research on Cancer/WHO, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer/WHO.

Funding

The authors acknowledge funding from the German Research Foundation (DFG 406710821) and the Agence Nationale de la Recherche; the Diet–Body–Brain (DietBB), the Competence Cluster in Nutrition Research (Federal Ministry of Education and Research, FKZ:01EA1410A); and the PerMiCCion project (Federal Ministry of Education and Research, Grant 100554612).

Data availability

Data described in the manuscript are not publicly available due to ethical restrictions, but are available upon reasonable request, pending application and approval. Data requests to be addressed to Prof. Ute Nöthlings via epi@uni-bonn.de.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tjnut.2024.09.023.

References

[1] C.M. Weaver, J. Dwyer, V.L. Fulgoni III, J.C. King, G.A. Leveille, R.S. MacDonald, et al., Processed foods: contributions to nutrition, Am. J. Clin. Nutr. 99 (2014) 1525–1542.
[2] C.A. Monteiro, G. Cannon, R.B. Levy, J.C. Moubarac, M.L. Louzada, F. Rauber, et al., Ultra-processed foods: what they are and how to identify them, Public Health Nutr 22 (2019) 936–941.
[3] F. Juul, A.L. Deierlein, G. Vaidean, P.A. Quatromoni, N. Parekh, Ultra-processed foods and cardiometabolic health outcomes: from evidence to practice, Curr. Atheroscler. Rep. 24 (2022) 849–860.
[4] M.M. Lane, E. Gamage, S. Du, D.N. Ashtree, A.J. McGuinness, S. Gauci, et al., Ultra-processed food exposure and adverse health outcomes: umbrella review of epidemiological meta-analyses, BMJ 384 (2024) e077310.
[5] H. Li, Y. Wang, E. Sonestedt, Y. Borné, Associations of ultra-processed food consumption, circulating protein biomarkers, and risk of cardiovascular disease, BMC Med 21 (2023) 415.
[6] L.G. Baraldi, E. Martinez Steele, D.S. Canella, C.A. Monteiro, Consumption of ultra-processed foods and associated sociodemographic factors in the USA between 2007 and 2012: evidence from a nationally representative cross-sectional study, BMJ Open 8 (2018) e020574.
[7] E. Mertens, C. Colizzi, J.L. Peñalvo, Ultra-processed food consumption in adults across Europe, Eur. J. Nutr. 61 (2022) 1521–1539.
[8] S.J. Dicken, R.L. Batterham, Ultra-processed food and obesity: what is the evidence? Curr. Nutr. Rep. 13 (2024) 23–38.
[9] R. Cordova, V. Viallon, E. Fontvieille, L. Peruchet-Noray, A. Jansana, K.H. Wagner, et al., Consumption of ultra-processed foods and risk of multimorbidity of cancer and cardiometabolic diseases: a multinational cohort study, Lancet Reg. Health Eur. 35 (2023) 100771.
[10] D. Su, J. Chen, S. Du, H. Kim, B. Yu, K.E. Wong, et al., Metabolomic markers of ultra-processed food and incident CKD, Clin. J. Am. Soc. Nephrol. 18 (2023) 327–336.
[11] K. Chang, M.J. Gunter, F. Rauber, R.B. Levy, I. Huybrechts, N. Kliemann, et al., Ultra-processed food consumption, cancer risk and cancer mortality: a large-scale prospective analysis within the UK Biobank, EClinicalMedicine 56 (2023) 101840.
[12] S. Wu, Z. Yang, S. Liu, Q. Zhang, S. Zhang, S. Zhu, Ultra-processed food consumption and long-term risk of irritable bowel syndrome: a large-scale prospective cohort study, Clin. Gastroenterol. Hepatol. 22 (2024) 1497–1507.e5.
[13] C. Samuthpongtorn, L.H. Nguyen, O.I. Okereke, D.D. Wang, M. Song, A.T. Chan, et al., Consumption of ultraprocessed food and risk of depression, JAMA Netw, Open 6 (2023) e2334770.
[14] B.M. Popkin, D.R. Miles, L.S. Taillie, E.K. Dunford, A policy approach to identifying food and beverage products that are ultra-processed and high in added salt, sugar and saturated fat in the United States: a cross-sectional analysis of packaged foods, Lancet Reg. Health Am. 32 (2024) 100713.
[15] D. Knorr, Food processing: legacy, significance and challenges, Trends Food Sci. Technol. 143 (2024) 104270.
[16] C.A. Monteiro, A. Astrup, Does the concept of "ultra-processed foods" help inform dietary guidelines, beyond conventional classification systems? YES, Am. J. Clin. Nutr. 116 (2022) 1476–1481.
[17] R.H. Lustig, Ultraprocessed food: addictive, toxic, and ready for regulation, Nutrients 12 (2020) 3401.
[18] A. Astrup, C.A. Monteiro, D.S. Ludwig, Does the concept of "ultra-processed foods" help inform dietary guidelines, beyond conventional classification systems? NO, Am. J. Clin. Nutr. 116 (2022) 1482–1488.
[19] J.M. Hess, M.E. Comeau, S. Casperson, J.L. Slavin, G.H. Johnson, M. Messina, et al., Dietary guidelines meet NOVA: developing a menu for a healthy dietary pattern using ultra-processed foods, J. Nutr. 153 (2023) 2472–2481.
[20] E.P. Ryan, A.L. Heuberger, C.D. Broeckling, E.C. Borresen, C. Tillotson, J.E. Prenni, Advances in nutritional metabolomics, Curr. Metab 1 (2013) 109–120.
[21] B. Zhou, R. Ichikawa, L.D. Parnell, S.E. Noel, X. Zhang, S.N. Bhupathiraju, et al., Metabolomic links between sugar-sweetened beverage intake and obesity, J. Obes. 2020 (2020) 7154738.
[22] A.C. Wood, G. Graca, M. Gadgil, M.K. Senn, M.A. Allison, I. Tzoulaki, et al., Untargeted metabolomic analysis investigating links between unprocessed red meat intake and markers of inflammation, Am. J. Clin. Nutr. 118 (2023) 989–999.
[23] X. Dong, Z. Zhuang, Y. Zhao, Z. Song, W. Xiao, W. Wang, et al., Unprocessed red meat and processed meat consumption, plasma metabolome, and risk of ischemic heart disease: a prospective cohort study of UK Biobank, J. Am. Heart Assoc. 12 (2023) e027934.
[24] T.M. McKeever, S.A. Lewis, P.A. Cassano, M. Ocké, P. Burney, J. Britton, et al., Patterns of dietary intake and relation to respiratory disease, forced expiratory volume in 1 s, and decline in 5-y forced expiratory volume, Am. J. Clin. Nutr. 92 (2010) 408–415.
[25] S. Andraos, M. Wake, R. Saffery, D. Burgner, M. Kussmann, J. O'Sullivan, Perspective: advancing understanding of population nutrient-health relations via metabolomics and precision phenotypes, Adv. Nutr. 10 (2019) 944–952.
[26] M.B. Schulze, M.A. Martínez-González, T.T. Fung, A.H. Lichtenstein, N.G. Forouhi, Food based dietary patterns and chronic disease prevention, BMJ 361 (2018) k2396.
[27] S. Andraos, K.L. Beck, M.B. Jones, T.L. Han, C.A. Conlon, J.V. de Seymour, Characterizing patterns of dietary exposure using metabolomic profiles of human biospecimens: a systematic review, Nutr. Rev. 80 (2022) 699–708.
[28] L.E. O'Connor, K.D. Hall, K.A. Herrick, J. Reedy, S.T. Chung, M. Stagliano, et al., Metabolomic profiling of an ultraprocessed dietary pattern in a domiciled randomized controlled crossover feeding trial, J. Nutr. 153 (2023) 2181–2192.
[29] I. Huybrechts, F. Rauber, G. Nicolas, C. Casagrande, N. Kliemann, R. Wedekind, et al., Characterization of the degree of food processing in the European Prospective Investigation into Cancer and Nutrition: application of the NOVA classification and validation using selected biomarkers of food processing, Front. Nutr. 9 (2022) 1035580.
[30] E. Handakas, K. Chang, N. Khandpur, E.P. Vamos, C. Millett, F. Sassi, et al., Metabolic profiles of ultra-processed food consumption and their role in obesity risk in British children, Clin. Nutr. 41 (2022) 2537–2548.
[31] N. Stratakis, A.P. Siskos, E. Papadopoulou, A.N. Nguyen, Y. Zhao, K. Margetaki, et al., Urinary metabolic biomarkers of diet quality in European children are associated with metabolic health, eLife 11 (2022) e71332.
[32] I. Perrar, U. Alexy, U. Nöthlings, Cohort profile update-overview of over 35 years of research in the Dortmund Nutritional and Anthropometric

Longitudinally Designed (DONALD) study, Eur. J. Nutr. 63 (2024) 727–740.
[33] Samuel Muli, Maike E Schnermann, Mira Merdas, Jodi Rattner, David Achaintre, Ines Perrar, Jantje Goerdten, Ute Alexy, Augustin Scalbert, Matthias Schmid, Anna Floegel, Pekka Keski-Rahkonen, Kolade Oluwagbemigun, Ute Nöthlings, Metabolomics signatures of sweetened beverages and added sugar are related to anthropometric measures of adiposity in young individuals: results from a cohort study, Am. J. Clin. Nutr. 120 (4) (2024) 879–890, https://doi.org/10.1016/j.ajcnut.2024.07.021.
[34] W. Sichert-Hellert, M. Kersting, C. Chahda, R. Schäfer, A. Kroke, German food composition database for dietary evaluations in children and adolescents, J. Food Compost. Anal. 20 (2007) 63–70.
[35] Bundesministerium für Ernährung und Landwirtschaft, Leitsätze für Gewürze und andere würzende Zutaten (Neufassung). https://www.bmel.de/SharedDocs/Downloads/DE/_Ernaehrung/Lebensmittel-Kennzeichnung/LeitsaetzeGewuerze.html, 1998.
[36] Y. Li, Y. Lai, T. Geng, P.F. Xia, J.X. Chen, Z.Z. Tu, et al., Association of ultraprocessed food consumption with risk of cardiovascular disease among individuals with type 2 diabetes: findings from the UK Biobank, Mol. Nutr. Food Res. 68 (2024) e2300314.
[37] B. Srour, L.K. Fezeu, E. Kesse-Guyot, B. Allès, C. Méjean, R.M. Andrianasolo, et al., Ultra-processed food intake and risk of cardiovascular disease: prospective cohort study (NutriNet-Santé), BMJ 365 (2019) l1451.
[38] M.L. Booth, A.D. Okely, T.N. Chey, A. Bauman, The reliability and validity of the Adolescent Physical Activity Recall Questionnaire, Med. Sci. Sports Exerc. 34 (2002) 1986–1995.
[39] A.M. Evans, C.D. DeHaven, T. Barrett, M. Mitchell, E. Milgram, Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems, Anal. Chem. 81 (2009) 6656–6667.
[40] Y. Yuan, J. Du, J. Luo, Y. Zhu, Q. Huang, M. Zhang, Discrimination of missing data types in metabolomics data based on particle swarm optimization algorithm and XGBoost model, Sci. Rep. 14 (2024) 152.
[41] S. Bijlsma, I. Bobeldijk, E.R. Verheij, R. Ramaker, S. Kochhar, I.A. Macdonald, et al., Large-scale human metabolomics studies: a strategy for data (pre-) processing and validation, Anal. Chem. 78 (2006) 567–574.
[42] R. Wei, J. Wang, M. Su, E. Jia, S. Chen, T. Chen, et al., Missing value imputation approach for mass spectrometry-based metabolomics data, Sci. Rep. 8 (2018) 663.
[43] N. Bararpour, F. Gilardi, C. Carmeli, J. Sidibe, J. Ivanisevic, T. Caputo, et al., DBnorm as an R package for the comparison and selection of appropriate statistical methods for batch effect correction in metabolomic studies, Sci. Rep. 11 (2021) 5657.
[44] N.B. Erichson, P. Zeng, K. Manohar, S.L. Brunton, J.N. Kutz, A.Y. Aravkin, Sparse principal component analysis via variable projection, SIAM J. Appl. Math. 80 (2020) 977–1002.
[45] G.R. Goldberg, A.E. Black, S.A. Jebb, T.J. Cole, P.R. Murgatroyd, W.A. Coward, et al., Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording, Eur. J. Clin. Nutr. 45 (1991) 569–581.
[46] L.O. Dragsted, Q. Gao, A. Scalbert, G. Vergères, M. Kolehmainen, C. Manach, et al., Validation of biomarkers of food intake-critical assessment of candidate biomarkers, Genes Nutr 13 (2018) 14.
[47] C. Cuparencu, T. Bulmuş-Tüccar, J. Stanstrup, G. La Barbera, H. Roager, L.O. Dragsted, Towards nutrition with precision: unlocking biomarkers as dietary assessment tools, Nat. Metab. 6 (8) (2024) 1438–1453, https://doi.org/10.1038/s42255-024-01067-y.
[48] G. Yang, S. Ge, R. Singh, S. Basu, K. Shatzer, M. Zen, et al., Glucuronidation: driving factors and their impact on glucuronide disposition, Drug Metab. Rev. 49 (2017) 105–138.
[49] T. Kyaw, P.J. Turnbaugh, Tiny gatekeepers: microbial control of host drug transporters, Clin. Pharmacol. Ther. 112 (2022) 443–445.
[50] P. Abiega-Franyutti, V. Freyre-Fonseca, Chronic consumption of food-additives lead to changes via microbiota gut-brain axis, Toxicology 464 (2021) 153001.
[51] Z. Song, R. Song, Y. Liu, Z. Wu, X. Zhang, Effects of ultra-processed foods on the microbiota-gut-brain axis: the bread-and-butter issue, Food Res. Int. 167 (2023) 112730.
[52] G.G. Kang, N.L. Trevaskis, A.J. Murphy, M.A. Febbraio, Diet-induced gut dysbiosis and inflammation: key drivers of obesity-driven NASH, iScience 26 (2022) 105905.
[53] C.R. Brydges, O. Fiehn, H.S. Mayberg, H. Schreiber, S.M. Dehkordi, S. Bhattacharyya, et al., Indoxyl sulfate, a gut microbiome-derived uremic toxin, is associated with psychic anxiety and its functional magnetic resonance imaging-based neurologic signature, Sci. Rep. 11 (2021) 21011.
[54] Y.C. Yeh, M.F. Huang, S.S. Liang, S.J. Hwang, J.C. Tsai, T.L. Liu, et al., Indoxyl sulfate, not p-cresyl sulfate, is associated with cognitive impairment in early-stage chronic kidney disease, Neurotoxicology 53 (2016) 148–152.
[55] S. Adesso, T. Magnus, S. Cuzzocrea, M. Campolo, B. Rissiek, O. Paciello, et al., Indoxyl sulfate affects glial function increasing oxidative stress and neuroinflammation in chronic kidney disease: Interaction between astrocytes and microglia, Front. Pharmacol. 8 (2017) 370.
[56] V. Shankar, M. Gouda, J. Moncivaiz, A. Gordon, N.V. Reo, L. Hussein, et al., Differences in gut metabolites and microbial composition and functions between Egyptian and U.S. children are consistent with their diets, mSystems 2 (2017) e00169, 16.
[57] EFSA Panel on Biological Hazards (BIOHAZ), Scientific opinion on risk based control of biogenic amine formation in fermented foods, EFSA J 9 (2011) 2393.
[58] S. Sánchez-Pérez, R. Celorio-Sardà, M.T. Veciana-Nogués, M.L. Latorre-Moratalla, O. Comas-Basté, M.C. Vidal-Carou, 1-methylhistamine as a potential biomarker of food histamine intolerance. A pilot study, Front. Nutri. 9 (2022) 973682.
[59] J.L. OrdÓÑez, R. CallejÓn, Biogenic amines in non-fermented food, in: B. Saad, R. Tofalo (Eds.), Biogenic Amines in Food: Analysis, Occurrence and Toxicity, the Royal Society of Chemistry, 2019, pp. 76–102, ch. 5, https://books.rsc.org/books/edited-volume/796/chapter-abstract/535510/Biogenic-Amines-in-Non-fermented-Food?redirectedFrom=fulltext.
[60] A.P. Doumatey, D. Shriner, J. Zhou, L. Lei, G. Chen, O. Oluwasola-Taiwo, et al., Untargeted metabolomic profiling reveals molecular signatures associated with type 2 diabetes in Nigerians, Genome Med 16 (2024) 38.
[61] M. Ulaszewska, N. Vázquez-Manjarrez, M. Garcia-Aloy, R. Llorach, F. Mattivi, L.O. Dragsted, et al., Food intake biomarkers for apple, pear, and stone fruit, Genes Nutr 13 (2018) 29.
[62] Y. Wang, S.M. Gapstur, B.D. Carter, T.J. Hartman, V.L. Stevens, M.M. Gaudet, et al., Untargeted metabolomics identifies novel potential biomarkers of habitual food intake in a cross-sectional study of postmenopausal women, J. Nutr. 148 (2018) 932–943.
[63] D.C. Nieman, N.D. Gillitt, W. Sha, D. Esposito, S. Ramamoorthy, Metabolic recovery from heavy exertion following banana compared to sugar beverage or water only ingestion: a randomized, crossover trial, PLOS ONE 13 (2018) e0194843.
[64] Y. Wang, R.A. Hodge, V.L. Stevens, T.J. Hartman, M.L. McCullough, Identification and reproducibility of plasma metabolomic biomarkers of habitual food intake in a US diet validation study, Metabolites 10 (2020) 382.
[65] W. Wagner, K. Sobierajska, Ł. Pułaski, A. Stasiak, W.M. Ciszewski, Whole grain metabolite 3,5-dihydroxybenzoic acid is a beneficial nutritional molecule with the feature of a double-edged sword in human health: a critical review and dietary considerations, Crit. Rev. Food Sci. Nutr. (2023) 1–19.
[66] L.H. Münger, A. Trimigno, G. Picone, C. Freiburghaus, G. Pimentel, K.J. Burton, et al., 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. 16 (2017) 3321–3335.
[67] N. Vázquez-Manjarrez, M. Ulaszewska, M. Garcia-Aloy, F. Mattivi, G. Praticò, L.O. Dragsted, et al., Biomarkers of intake for tropical fruits, Genes Nutr 15 (2020) 11.
[68] N.E. Allen, P.B. Grace, A. Ginn, R.C. Travis, A.W. Roddam, P.N. Appleby, et al., Phytanic acid: measurement of plasma concentrations by gas-liquid chromatography-mass spectrometry analysis and associations with diet and other plasma fatty acids, Br. J. Nutr. 99 (2008) 653–659.
[69] J. Kaur, M. Gulati, S.K. Singh, G. Kuppusamy, B. Kapoor, V. Mishra, et al., Discovering multifaceted role of vanillic acid beyond flavours: nutraceutical and therapeutic potential, Trends Food Sci. Technol. 122 (2022) 187–200.
[70] EFSA Panel on Food Additives and Flavourings (FAF), M. Younes, G. Aquilina, L. Castle, G. Degen, K.-H. Engel, et al., Flavouring

group evaluation 413 (FGE.413): naringenin, EFSA J 22 (2024) e8747.
[71] Y. Yang, M. Trevethan, S. Wang, L. Zhao, Beneficial effects of citrus flavanones naringin and naringenin and their food sources on lipid metabolism: an update on bioavailability, pharmacokinetics, and mechanisms, J. Nutr. Biochem. 104 (2022) 108967.
[72] M. Yadav, A. Lomash, S. Kapoor, R. Pandey, N.S. Chauhan, Mapping of the benzoate metabolism by human gut microbiome indicates food-derived metagenome evolution, Sci. Rep. 11 (2021) 5561.
[73] M. Gelpi, F. Mikaeloff, A.D. Knudsen, R. Benfeitas, S. Krishnan, S Svenssson Akusjärvi, et al., The central role of the glutamate metabolism in long-term antiretroviral treated HIV-infected individuals with metabolic syndrome, Aging (Albany NY) 13 (2021) 22732–22751.
[74] U. Sovio, N. McBride, A.M. Wood, K.L. Masconi, E. Cook, F. Gaccioli, et al., 4-Hydroxyglutamate is a novel predictor of pre-eclampsia, Int. J. Epidemiol. 49 (2020) 301–311.
[75] X. Zhao, Y. Wang, L. Li, J. Mei, X. Zhang, Z. Wu, Predictive value of 4-Hydroxyglutamate and miR-149-5p on eclampsia, Exp. Mol. Pathol. 119 (2021) 104618.
[76] S.P. Claus, H. Guillou, S. Ellero-Simatos, The gut microbiota: a major player in the toxicity of environmental pollutants? NPJ Biofilms Microbiomes 2 (2016) 16003.
[77] K.B. Martinez, V. Leone, E.B. Chang, Western diets, gut dysbiosis, and metabolic diseases: are they linked? Gut Microbes 8 (2017) 130–142.

Extended analyses on publication 4: UPF related metabolites associated with adiposity in adolescents and young adults using penalized elastic net regression models

Biosample	**Metabolite**	**BMI** (*n* = 339)[1]	**% BF** (*n* = 339)[1]	**WC** (*n* = 231)[1]
Adolescent Urine	N,N-dimethylalanine	-0.125	-0.144	-0.188
	X - 17358	0.100	0.041	0.069
	4-hydroxymandelate	-0.097	-0.118	-0.078
	2-acetamidophenol sulfate	-0.070	-0.069	-0.107
	dopamine 3-O-sulfate	0.069	—	0.042
	3-methyladipate	-0.067	—	-0.001
	X - 17679	-0.060	—	-0.087
	X - 12818	-0.058	-0.029	—
	X - 13844	-0.056	-0.073	-0.056
	glucuronide of C10H18O2 (9)*	0.044	0.033	0.022
	2S,3R-dihydroxybutyrate	0.042	—	—
	3,5-dihydroxybenzoic acid	-0.040	—	—
	4-hydroxycinnamate sulfate	-0.029	-0.039	-0.029
	glycerophosphorylcholine (GPC)	-0.027	—	-0.020
	4-methoxyphenol sulfate	0.026	—	—
	picolinoylglycine	0.016	—	—
	X - 24345	0.012	—	0.070
	heptenedioate (C7:1-DC)*	0.004	—	—
	3-indoxyl sulfate	-0.001	—	—
	X - 17825	—	0.027	—
	glycolate (hydroxyacetate)	—	0.010	0.063
	hydroquinone sulfate	—	—	-0.032
Young adult Plasma		BMI (*n* = 195)[1]	% BF (*n* = 195)[1]	WC (*n* = 195)[1]
	4-hydroxyglutamate	0.304	0.230	0.318
	X - 11372	-0.149	-0.126	-0.177
	X - 23639	-0.042	—	-0.028
	3-carboxy-4-methyl-5-pentyl-2-furanpropionate (3-CMPFP)	0.039	—	0.140
	X - 24951	—	—	0.094

[1]Only UPF-related metabolites (in publication 4) that were also associated (non-zero values) with at least one of the three anthropometric measures are shown in this table. A dash [—] indicates that the metabolite had no association with the anthropometric measure.

Abbreviations: BMI, body mass index; %BF, body fat percentage; WC, waist circumference

Metabolites with a prefix 'X-' followed by a number (e.g., X - 17358) are molecular features whose biochemical identities could not be identified. *Indicates metabolites that were not confirmed based on authentic standard, but Metabolon Inc are confident in its identity based on their biochemical identification criteria and chemical properties.

The analytic approach for the adaptive elastic regression models and their nested cross-validation is reported in Publication 2 (and its supplemental material). Briefly, linear regression models were applied to adjust metabolite concentrations for confounders (age, sex, energy intake, birthweight, time difference between biosample collection and anthropometric measurements, physical activity, smoking, and alcohol status). For each of the adiposity measures (target Y variables, standardized to a mean of zero and unit variance), the confounder-adjusted metabolites (predictors) were fit in an adaptive elastic-net regression model, using the inverse of the absolute ridge regression weights as penalty factors, and a nested cross-validation approach implemented in the nestedcv (v0.4.4) R package.

4 Discussion with references

4.1 Main findings

In our systematic review (publication 1) we identified acesulfame, saccharin, sucralose, cyclamate, and steviol glucuronide as the most validated biomarkers for LNCSB intake. These were based on two experimental and one observational study using targeted metabolomics approaches. Applying untargeted approaches, we replicated acesulfame and saccharin as urinary biomarkers of LNCSB intake in children and adolescents in the DONALD cohort (publication 2). Acesulfame and saccharin, though biochemically different compounds, share certain characteristics regarding their absorption and metabolism, and are both primarily excreted in urine (Magnuson et al., 2016). Only saccharin was quantified in both children and adolescent urine samples. Acesulfame was detected in adolescent urine samples only. It was unclear whether this was partly due to inter-laboratory differences as children and adolescent samples were analyzed at different laboratories, or whether these results reflected differences in beverage choices and other dietary exposures across these two age groups. By using repeated diet and urine measurements in children, it was expected that even the transient dietary biomarkers that may be missed in single measurements would be quantified.

As exogenous compounds, urinary acesulfame and saccharin reflect dietary exposures to the parent sweetener. However, these sweeteners are common ingredients in other foods such as sweets and savory snacks. In our post-hoc analyses, there were no discernible correlations between these metabolites and other food groups. Other sweeteners were likely not observed in this study due to their different absorption, metabolism, and excretion pathways and were not quantified by our approaches; for example, stevio glycoside and aspartame are degraded into different compounds (Magnuson et al., 2016). Because of the sustained interest in these sweeteners, more advanced targeted approaches for simultaneously quantifying all the main industrial sweeteners in urine have been developed (Bruin et al., 2023), and applied in recent large population studies (Buso et al., 2024).

We highlighted the limitations of 24-h urinary sucrose and fructose for assessing SSB intake (publication 1). In line, we replicated 24-h urinary sucrose as a biomarker for AS in

adolescents, but not for SSB intake (publication 2). The other panel of SSB biomarkers in publication 1 (formate, citrulline, taurine, and isocitrate) had low validity based on the Dragsted framework for biomarkers of food intake (Dragsted et al., 2018), mainly due to lack of specificity for SSB intake and limited replication. Although formate, citrulline, and taurine were quantified in adolescent urine samples, they were not predictive of SSB intake. Strikingly, SSB intake was more positively related to caffeine metabolites, particularly, the 1-methylxanthine and 5- acetylamino-6-amino-3-methyluracil in adolescent urine and young adult plasma samples, independent of other dietary sources of caffeine (publication 2).

Caffeine is a major ingredient in soft drinks (e.g., sodas, energy drinks, and iced teas); hence, caffeine-related metabolites are biologically plausible ingredient-based biomarkers for caffeinated SSBs. Ingredient-based biomarkers are unlikely to be reliable in the assessment of heterogeneous foods such as SSB, as they do not generalize to subgroups not enriched with the primary ingredient. This could also explain our lack of replication of taurine, an ingredient-based biomarker of SSB, particularly, for cola drinks (Gibbons et al., 2015). Considering our previous findings on the carbon isotope ratio (δ13C) of alanine as the most validated biomarker of SSB as demonstrated by various targeted studies (publication 1), it is of interest to determine how the SSB metabolites from untargeted approaches (publication 2) correlate with the isotopic signature δ13C. Ultimately, combining various biomarkers appears to be a more promising strategy to advance the assessment of SSB intake.

Sweet and fatty snacks, like SSBs, are heterogeneous and among the most difficult foods to determine their accurate intakes due to the unplanned nature of snacking. Many are consumed at irregular intervals, in varying portion sizes, and are of diverse composition; hence, they are prone to misreporting (Garden et al., 2018). In our children analytic sample, three candidate biomarkers of chocolate intake were identified: theobromine, xanthosine, and cyclo(L-prolyl-L-valyl). These were externally replicated in an independent cohort (publication 3). The lack of reproducibility across metabolite features for other subgroups of snacks is likely due to their complex multi-ingredients and subgroup heterogeneity. Chocolate was the most homogenous subgroup of snacks included as the primary ingredient in most brands is cocoa beans. Besides, while theobromine is a known

biomarker for cocoa intake, it is also a metabolite of caffeine; thus, it lacks the specificity for chocolate (Michielsen et al., 2018).

In publication 4, our results showed that the complex nature of UPF and its effects on human metabolism are reflected in its associated metabolomic changes. A significant contribution of this study was the replication of many UPF-related metabolites previously reported in a controlled human feeding study (O'Connor et al., 2023), demonstrating the generalizability of these findings in a free-living population and their relevance to diverse UPF diets across populations. Our results showed that consumers of diets rich in UPF unsurprisingly consumed smaller amounts of whole, nutritious foods (e.g., vegetables, fruits, whole grains, legumes, etc.) that are the basis for diet quality and known to promote health. This was evident from the strong negative correlation between the UPF intake and unprocessed or minimally processed food intake, and from the inverse associations between UPF intake and metabolites known to reflect these whole foods (publication 4).

Even though the DONALD cohort generally includes healthy participants, some of the metabolites positively associated with UPF intake are validated markers for various metabolic health conditions. For example, 4-hydroxyglutamate has been linked to metabolic syndrome (Gelpi et al., 2021), primary hyperoxaluria type 3 (Pitt et al., 2015), pre-eclampsia (Sovio et al., 2019), and gestational diabetes mellitus (Sovio et al., 2022). The UPF-related metabolite Indoxyl sulfate is also linked to psychic anxiety (Brydges et al., 2021), cognitive impairment (Yeh et al., 2016), neuroinflammation and oxidative stress (Adesso et al., 2017). To gain a better understanding of the biological mechanisms underlying long-term consumption of UPFs and health outcomes, longitudinal studies should prioritize these metabolically and physiologically important metabolites.

Regarding diet-adiposity associations, the SBs-, AS-, and UPF-related metabolites that were also significantly associated with adiposity measures were interpreted as potential intermediate biomarkers of diet and adiposity, based on common metabolic pathway perturbations (Vineis et al., 2013). Overall, we discovered new associations as well as replicated associations reported in literature. For example, N1-methyl-2-pyridone-5-carboxamide and decanoylcarnitine (C10), both positively related to all three measures of adiposity in adolescents (publication 2), were also positively associated with adiposity in

other cohorts (Goodson et al., 2019; Lowe et al., 2017), respectively. Nonetheless, these results were observed in a single urine measurement, and longitudinal studies would provide a better interpretation.

In young adult plasma samples, which assessed long-term dietary intakes preceding the blood draw, two results were particularly noteworthy. First, higher SSB consumers had lower concentrations of plasma carotene diol (publication 2), one of the carotenoid markers of leafy vegetable intakes (Landberg et al., 2023). Also, the young adults with higher WC had lower levels of carotene diol (publication 2). Other epidemiologic studies have reported lower levels of serum carotene diol in individuals with higher BMI and WC (Stevens et al., 2020) and similarly, other carotenoids with visceral adiposity (Yan et al., 2023). Carotenoids are major phytochemicals in fruits and vegetables and are indicative of adherence to healthier dietary patterns (Neuhouser et al., 2023; Holthaus et al., 2024). Collectively, our findings suggest a potential indirect relationship between SSB and adiposity, but more plausibly through overall quality of diet, since higher SSB consumers also tend to consume other foods of poorer diet quality (Doherty, Lacko, & Popkin, 2021).

Secondly, plasma 4-hydroxyglutamate which was positively related with all three measures of adiposity in young adults, has been reported to mediate the effects of maternal overweight and obesity on early childhood growth trajectories and obesity risk (Hu et al., 2022). As a primary metabolite, 4-hydroxyglutamate is a metabolically and physiologically important metabolite considering its other associations with metabolic health conditions. Mechanistic studies should provide a better understanding of the specific mechanisms through which UPF, or its subgroups contribute to higher levels of 4-hydroxyglutamate, and its functional roles in cellular metabolism and pathophysiology of related health conditions.

4.2 Reflections and conclusion

Metabolomics has significantly accelerated progress towards precision nutrition (Landberg et al., 2023; Prentice, 2024; Cuparencu et al., 2024), but there are still some unresolved issues regarding the most optimal approaches for biomarker research. For instance, nearly all metabolomics-based studies in free-living populations, including the present study, rely on self-reported intakes for identification of biomarkers of intake. This

appears to be a counterintuitive and suboptimal strategy for advancing dietary biomarker research, as the need for intake biomarkers arises from the uncertainty and errors in these self-reported intakes (Prentice, 2024). Today, we have an ever-growing, extensive list of candidate dietary biomarkers, many of which have not been replicated across studies or validated in intervention studies. Our publication 3, for example, demonstrates the challenges of using untargeted approaches for identification and replication of biomarkers of intake in free-living populations. Ultimately, there is a need to strike a balance between the pursuit of novelty provided by these untargeted approaches and validation of existing candidate biomarkers of intake. As evident across publications 1-4, human controlled feeding studies appear more suited for food biomarker discovery, and observational studies for evaluating their generalizability in real-world settings. Targeted metabolomics seems more ideal for the latter.

In conclusion, we demonstrated the usefulness of the human metabolome in elucidating the mechanisms through which diet influences metabolism and adiposity. These findings only fit a piece of the puzzle on the diet-adiposity relationship. Integrating metabolome with other omics (e.g., microbiome, proteomics, and genomics) can advance this work, particularly in longitudinal studies with better defined temporality of dietary exposures, omics data, and future adiposity measurements. To address some of the limitations discussed in publications 2-4, we intend to extend this work by applying a life-course approach to dietary intakes where the same individuals are investigated from early childhood to adulthood to gain an even better understanding of the effects of these foods on microbiome composition and some of the physiologically important gut-microbial metabolites associated with adiposity and other metabolic health conditions.

4.3 References

Adesso S, Magnus T, Cuzzocrea S, Campolo M, Rissiek B, Paciello O, et al. Indoxyl sulfate affects glial function increasing oxidative stress and neuroinflammation in chronic kidney disease: interaction between astrocytes and microglia. Front Pharmacol 2017; 8: 370

Bruin M, Maho W, Buso MEC, Naomi ND, Brouwer-Brolsma EM, Feskens EJM, et al. Development and validation of a UPLC-MS/MS method for the quantification of sugars and non-nutritive sweeteners in human urine. J Chromatogr B Analyt Technol Biomed Life Sci 2023; 1225: 123741

Brydges CR, Fiehn O, Mayberg HS, Schreiber H, Dehkordi S, Bhattacharyya S, et al. Indoxyl sulfate, a gut microbiome-derived uremic toxin, is associated with psychic anxiety and its functional magnetic resonance imaging-based neurologic signature. Sci Rep 2021; 11: 21011

Buso M, Boshuizen HC, Naomi ND, Maho W, Bruin M, Balvers M, et al. Relative validity of habitual sugar and low/no-calorie sweetener consumption assessed by food frequency questionnaire, multiple 24-h dietary recalls and urinary biomarkers: an observational study within the SWEET project. Am J Clin Nutr 2024; 119: 546-559

Cuparencu C, Bulmuş-Tüccar T, Stanstrup J, La Barbera G, Roager HM, Dragsted LO. Towards nutrition with precision: unlocking biomarkers as dietary assessment tools. Nat Metab 2024; 6: 1438-1453

Doherty A, Lacko A, Popkin B. Sugar-sweetened beverage (SSB) consumption is associated with lower quality of the non-SSB diet in US adolescents and young adults. Am J Clin Nutr 2021; 113: 657-664

Dragsted LO, Gao Q, Scalbert A, Vergères G, Kolehmainen M, Manach C, et al. Validation of biomarkers of food intake—critical assessment of candidate biomarkers. Genes Nutr 2018; 13:14

Garden L, Clark H, Whybrow S, Stubbs RJ. Is misreporting of dietary intake by weighed food records or 24-hour recalls food specific? Eur J Clin Nutr 2018; 72: 1026-1034

Gelpi M, Mikaeloff F, Knudsen AD, Benfeitas R, Krishnan S, Svenssson Akusj, et al. The central role of the glutamate metabolism in long-term antiretroviral treated HIV-infected individuals with metabolic syndrome. Aging (Albany NY) 2021; 13: 22732-22751

Gibbons H, McNulty BA, Nugent AP, Walton J, Flynn A, Gibney MJ, et al. A metabolomics approach to the identification of biomarkers of sugar-sweetened beverage intake. Am J Clin Nutr 2015; 101: 471-477

Goodson JM, Hardt M, Hartman ML, Alqaderi H, Green D, Tavares M, et al. Salivary N1-Methyl-2-Pyridone-5-Carboxamide, a biomarker for uranium uptake, in Kuwaiti children exhibiting exceptional weight gain. Front Endocrinol (Lausanne). 2019; 10: 382

Holthaus TA, Keye SA, Verma S, Cannavale CN, Burd NA, Holscher HD, et al. Dietary patterns and carotenoid intake: comparisons of MIND, Mediterranean, DASH, and Healthy Eating Index. Nutr Res 2024; 126: 58-66

Hu Z, Han L, Liu J, Fowke JH, Han JC, Kakhniashvili D, et al. Prenatal metabolomic profiles mediate the effect of maternal obesity on early childhood growth trajectories and obesity risk: the Conditions Affecting Neurocognitive Development and Learning in Early Childhood (CANDLE) Study. Am J Clin Nutr 2022; 116: 1343-1353

Landberg R, Karra P, Hoobler R, Loftfield E, Huybrechts I, Rattner JI, et al. Dietary biomarkers—an update on their validity and applicability in epidemiological studies. Nutr Rev 2023; 82: 1260-1280

Lowe WL Jr, Bain JR, Nodzenski M, Reisetter AC, Muehlbauer MJ, Stevens RD, et al. Maternal BMI and glycemia impact the fetal metabolome. Diabetes Care 2017; 40: 902-910

Magnuson BA, Carakostas MC, Moore NH, Poulos SP, Renwick AG. Biological fate of low-calorie sweeteners. Nutr Rev 2016; 74: 670-689

Michielsen C, Almanza-Aguilera E, Brouwer-Brolsma EM, Urpi-Sarda M, Afman LA. Biomarkers of food intake for cocoa and liquorice (products): a systematic review. Genes Nutr 2018; 13: 22

Neuhouser ML, Prentice RL, Tinker LF, Lampe JW. Enhancing capacity for food and nutrient intake assessment in population sciences research. Annu Rev Public Health 2023; 44: 37-54

O'Connor LE, Hall KD, Herrick KA, Reedy J, Chung ST, Stagliano M, et al. Metabolomic profiling of an ultraprocessed dietary pattern in a domiciled randomized controlled crossover feeding trial. J Nutr 2023; 153: 2181-2192

Pitt JJ, Willis F, Tzanakos N, Belostotsky R, Frishberg Y. 4-Hydroxyglutamate is a biomarker for primary hyperoxaluria type 3. JIMD Rep 2015; 15: 1-6

Prentice RL. Intake biomarkers for nutrition and health: review and discussion of methodology issues. Metabolites 2024; 14: 276

Sovio U, Clayton GL, Cook E, Gaccioli F, Charnock-Jones DS, Lawlor DA, et al. Metabolomic identification of a novel, externally validated predictive test for gestational diabetes mellitus. J Clin Endocrinol Metab 2022; 107: e3479-e86

Sovio U, McBride N, Wood AM, Masconi KL, Cook E, Gaccioli F, et al. 4-Hydroxyglutamate is a novel predictor of pre-eclampsia. Int J Epidemiol 2019; 49: 301-311

Stevens VL, Carter BD, McCullough ML, Campbell PT, Wang Y. Metabolomic profiles associated with BMI, waist circumference, and diabetes and inflammation biomarkers in women. Obesity (Silver Spring) 2020; 28: 187-196

Vineis P, van Veldhoven K, Chadeau-Hyam M, Athersuch TJ. Advancing the application of omics-based biomarkers in environmental epidemiology. Environ Mol Mutagen 2013; 54: 461-467

Yan S, Chen S, Liu Y, Liang H, Zhang X, Zhang Q, et al. Associations of serum carotenoids with visceral adiposity index and lipid accumulation product: a cross-sectional study based on NHANES 2001–2006. Lipids Health Dis 2023; 22: 209

Yeh YC, Huang MF, Liang SS, Hwang SJ, Tsai JC, Liu TL, et al. Indoxyl sulfate, not p-cresyl sulfate, is associated with cognitive impairment in early-stage chronic kidney disease. Neurotoxicology. 2016;53:148-52

5. Acknowledgements

I am grateful to my PhD advisor, Prof. Ute Nöthlings, for her guidance and mentorship throughout my doctoral studies. I also thank my second advisor, Prof. Matthias Schmid, for sharing his statistical expertise with me and for his constructive feedback on my work. Many thanks also go to the other members of my thesis advisory committee, Prof. Markus Nöthen and Jun-Prof. Marie-Christine Simon, for walking this journey with me.

A heartfelt shout-out to Dr. Dr. Kolade Oluwagbemigun, the postdoctoral epidemiologist I was incredibly fortunate to work closely with throughout my doctoral research. For the mentorship and support, I got nothing but a deep, enduring gratitude.

Special thanks to my family. Even when many miles away, I knew you were there, cheering me on. This is for you!