

Dietary biomarkers—an update on their validity and applicability in epidemiological studies

Rikard Landberg ^{1,*}, Prasoon Karra^{2,3}, Rachel Hoobler^{2,3}, Erikka Loftfield ⁴, Inge Huybrechts ⁵, Jodi I. Rattner⁵, Stefania Noerman¹, Liesel Claeys⁶, Vanessa Neveu⁵, Nanna Hjort Vidkjaer ¹, Otto Savolainen¹, Mary C. Playdon^{2,3}, and Augustin Scalbert⁵

¹Division of Food and Nutrition Science, Department of Life Sciences, Chalmers University of Technology, Gothenburg, Sweden

²Department of Nutrition and Integrative Physiology, University of Utah, Salt Lake City, UT, USA

³Cancer Control and Population Sciences Program, Huntsman Cancer Institute, University of Utah Salt Lake City, UT, USA

⁴Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA

⁵International Agency for Research on Cancer, Nutrition and Metabolism Branch, Lyon, France

⁶International Agency for Research on Cancer, Molecular Mechanisms and Biomarkers Group, Lyon, France

*Correspondence: R. Landberg, Division of Food and Nutrition Science, Department of Life Sciences, Chalmers University of Technology, Kemigården 4, Gothenburg 412 96, Sweden. E-mail: Rikard.landberg@chalmers.se.

The aim of this literature review was to identify and provide a summary update on the validity and applicability of the most promising dietary biomarkers reflecting the intake of important foods in the Western diet for application in epidemiological studies. Many dietary biomarker candidates, reflecting intake of common foods and their specific constituents, have been discovered from intervention and observational studies in humans, but few have been validated. The literature search was targeted for biomarker candidates previously reported to reflect intakes of specific food groups or components that are of major importance in health and disease. Their validity was evaluated according to 8 predefined validation criteria and adapted to epidemiological studies; we summarized the findings and listed the most promising food intake biomarkers based on the evaluation. Biomarker candidates for alcohol, cereals, coffee, dairy, fats and oils, fruits, legumes, meat, seafood, sugar, tea, and vegetables were identified. Top candidates for all categories are specific to certain foods, have defined parent compounds, and their concentrations are unaffected by nonfood determinants. The correlations of candidate dietary biomarkers with habitual food intake were moderate to strong and their reproducibility over time ranged from low to high. For many biomarker candidates, critical information regarding dose response, correlation with habitual food intake, and reproducibility over time is yet unknown. The nutritional epidemiology field will benefit from the development of novel methods to combine single biomarkers to generate biomarker panels in combination with self-reported data. The most promising dietary biomarker candidates that reflect commonly consumed foods and food components for application in epidemiological studies were identified, and research required for their full validation was summarized.

Key words: blood, cohort, dietary biomarkers, epidemiology, food component biomarkers, food intake biomarkers, urine, validation

INTRODUCTION

Diet is an important modifiable risk factor for noncommunicable diseases, including cardiovascular disease, type 2 diabetes, and certain cancers.¹ Evidence of dietary relationships with disease largely stems from observational studies, where self-reporting tools like food-frequency questionnaires (FFQs), 24-hour recalls (24-HRs), and weighed food records (FRs) have been used to estimate food intake.^{2,3} Yet, large random and systematic measurement errors hamper the accuracy of these tools to capture long-term food intake, although methods have been developed to tackle measurement errors, such as combining multiple 24-HRs.^{4,5} There is a need for new tools and methods to reflect long-term dietary exposures objectively and more accurately.^{5,6} Dietary biomarkers are promising instruments for objective dietary assessment, as they are molecules (molecular weight <1000 Da) derived from specific foods that are absorbed and detected in biological samples from humans in response to food intake, and they do not depend on participant recall, motivation, or behavior.⁷ Dietary biomarkers vary in their definition and applications. Recovery biomarkers provide a quantitative measure of food intake, as their excretion corresponds to the intake amount and can thus be used to correct for dietary measurement error. Concentration biomarkers correlate with food intake and can rank individuals with respect to food intake, although metabolism and other characteristics may affect their measured level. Replacement and prediction biomarkers are highly predictive of food intake, but they do not fulfill the requirements of recovery biomarkers.^{6,7} More recently, other classification schemes have emerged to account for other applications and features,^{5,8} including classifying exposure biomarkers into biomarkers of food component intake, food intake, and dietary patterns,⁸ either as single biomarkers or multiple-biomarker panels. Few valid and reliable food-related biomarkers exist at present, but recent developments in high-resolution mass spectrometry (MS), and to some extent nuclear magnetic resonance spectroscopy (NMR)-based techniques, have increasingly been used for diet-related biomarker discovery, validation, and implementation.^{9,10} Large databases now facilitate research and development of biomarker validation and implementation.^{11–13} Both controlled-feeding studies and large-scale epidemiological studies that leverage metabolomics have discovered novel biomarkers for a wide diversity of foods, food groups, and dietary patterns.^{14,15} Carefully controlled intervention studies are particularly useful to assess the pharmacokinetics of biomarker candidates as well as to establish dose response. Observational studies are useful to

characterize biomarker variability under free-living conditions, and to estimate long-term biomarker stability. In some cases, analyses of putative dietary biomarkers in relation to health outcomes have complemented those using self-reported dietary data.^{16–21} However, few candidate biomarkers meet all proposed criteria for validation, often because methodological studies are lacking. Although standardized validation schemes are scarce,²² typical validation criteria include the following: biomarker assay validation, biomarker kinetics (half-life), assessment of the correlation of the biomarker with true food intake (using surrogate measures), dose response, specificity and sensitivity, exploration of nonfood intake-related determinants, and between-person and within-person variation over time.^{7,23–25} Other criteria include robustness across studies of different designs.^{15,17,26}

The aim of this review was to identify and provide a summary update on the validity and applicability of the most promising dietary biomarkers reflecting important foods in the Western diet that can be applied in epidemiological studies. Each biomarker was evaluated against the recently developed 8-step validation process for diet-related biomarkers that systematically assesses candidate biomarker plausibility, dose response, time response, robustness, reliability, stability, analytical performance, and interlaboratory reproducibility.²⁵ This scheme was extended by evaluating data on biomarker reproducibility over time, and knowledge gaps and opportunities for future research are highlighted.

METHODS

Selected food items

The biomarker candidates searched for were previously reported to reflect intake of specific food groups or components that are of major importance in health and disease, or that are included in official dietary guidelines and by nongovernmental organizations such as the World Cancer Research Fund/American Institute for Cancer Research.²⁷ The Food Biomarker Alliance (FoodBALL; <https://www.wur.nl/en/project/foodball.htm>) extensively reviewed putative dietary biomarkers,²⁵ which served as the foundation for the current review. The literature on biomarkers of the following foods or food components was reviewed: alcohol, cereals, coffee, dairy, fats and oils, fish, fruits, legumes, meat, seafood, sugar, tea, and vegetables. For cereals, whole-grain wheat, rye, oats, as well as bran from wheat, rye, and sourdough fermented bread (rye) were included. Whole-grain and refined-grain rice, maize, millet, sorghum, barley, and their products were excluded given the lack of promising, specific candidate biomarkers for them.²⁶ Among other

plant foods, total fruit intake, total vegetable intake, specific fruits and vegetables, and legumes were reviewed. Intake biomarkers of processed meat, red meat, poultry,^{28,29} total fish, lean fish, and other seafood^{29,30} were evaluated. Putative dairy intake biomarkers included biomarkers of total dairy, milk, yogurt, and cheese products.^{31–34} Biomarkers of commonly consumed beverages like tea, coffee, and alcohol were also evaluated.³⁵ Among sugars, fructose and sucrose³⁶ were included in the review based on their extensive consumption.³⁷ Additionally, candidate biomarkers³⁸ of fat and oil intakes, as well for which the typical biomarkers were primarily determined by the fat or oil food source, were reviewed (eg, fish oil vs plant oils, etc).³⁶

Biomarker validation criteria

In this review, a modified version of the systematic framework for food intake biomarker validation defined and reported by the FoodBALL consortium²⁵ was used. The review focuses on biomarkers of habitual food intake in population studies and does not include biomarkers of compliance in short-term dietary intervention studies. In the FoodBALL consortium framework, 8 validation criteria that apply to different study designs include the following: plausibility (chemical/biological plausibility and specificity), dose response (across different levels of intake), time response (biomarker kinetics), robustness (reflection of a specific food in a whole-meal/diet context), reliability (comparable with other biomarkers or dietary instruments used to reflect the same food), stability (chemically and biologically), analytical performance (accuracy of the assay), and interlaboratory reproducibility (similar results across at least 2 laboratories).²⁵ In the modified criteria applicable to epidemiological studies (Table 1), the plausibility, dose–response, and time–response criteria from those defined by Dragsted et al²⁵ were included. Instead of robustness, a specific criterion for “correlation with habitual food intake” and “correlation with short-term food intake” was used, which addresses the correlation with intake under free-living conditions but does not formally consider that the biomarker must have been validated in controlled dietary intervention studies. The magnitude of correlation between human specimen-derived biomarkers and dietary intakes estimated by FFQs, 24-HRs, or FRs is represented by the correlation coefficient “*r*”. Correlations with $r < 0.2$ were considered “weak,” “moderate” when $r = 0.2$ – 0.5 , and “strong” when $r > 0.5$.³⁹ The stability criteria were excluded because biomarkers used from free-living studies typically rely on samples from biobanks and cohorts that have been stored over a longer time period and chemical stability tests of such storage conditions are lacking for

most biomarkers. The analytical performance criteria were simplified to indicate information regarding “biospecimen” and “analytical method” used to measure the biomarker. Finally, the interlaboratory reproducibility criterion was excluded because these data were largely unavailable. In addition, “reproducibility over time” was included, mainly represented by the intraclass correlation coefficient (ICC) of repeated measures over time to provide a measure of how well the long-term biomarker concentration could be reflected in a single measurement. Most cohorts provide biospecimens measured at a single time point and the candidate biomarkers are typically measured against habitual dietary intake during the prior 12 months. Reproducibility over time was considered as “poor” when $ICC < 0.4$, “fair” when $ICC = 0.4$ – 0.6 , “good” when $ICC = 0.60$ – 0.75 , and excellent when $ICC > 0.75$.⁴⁰

Selection of studies

Dietary candidate biomarkers have typically emerged from small short-term human feeding trials or cross-sectional population-based studies.⁵ To evaluate the different elements of dietary biomarker validity as outlined above, data primarily from cross-sectional studies nested within prospective cohort studies were included. Assessment of dose response used information from dietary intervention studies. Animal studies were not reviewed.

Search strategy and biomarker evaluation process

In this paper, the search strategies reported in the recent review articles on food intake biomarkers derived from the FoodBALL consortium²⁶ were replicated and extended. For each dietary exposure, we present a summary of candidate dietary biomarkers and validation criteria assessment (see Table S1 in the Supporting Information online). For each dietary exposure, standardized summary sheets of top biomarkers were compiled with appraisal of validation criteria along with key references (see Text S1 in the Supporting Information online).

RESULTS AND DISCUSSION

Table S1 in the Supporting Information online summarizes the results from a literature review of human epidemiological studies on candidate biomarkers of intake of specific foods, food groups, and food components. The largest number of candidate biomarkers were identified for intakes of cereals and beverages, and the least for biomarkers of dairy and legume intake. The most extensively validated candidate biomarkers are those

Table 1 Candidate dietary biomarker validation criteria

Candidate dietary biomarker validation criteria	Description
Nature of the biomarker and its specificity/plausibility	Is the biomarker a parent compound or a metabolite derived from intake/metabolism of the diet exposure? To what extent is the biomarker specific for the food?
Biospecimen	Is the biomarker present in plasma, urine, or other matrices (such as adipose tissue, nails, hair)?
Analytical method	What analytical methods were used to analyze the biomarker (LC, GC, NMR or other)?
Correlation with habitual food intake	What is the magnitude of correlation of the biomarker with habitual food intake assessed by FFQ?
Time response	What is the temporal relationship of the biomarker with food intake assessed by pharmacokinetics parameters (ie, elimination half-life)?
Reproducibility over time	What is the ratio of between-subject variation to the sum of between- and within-subject variation (ICC)? How does this relate to half-life and frequency of intake?
Dose response	What is the biomarker concentration following sequential increases in food intake under controlled or free-living conditions?

Abbreviations: FFQ, food-frequency questionnaire; FR, food record; GC, gas chromatography; ICC, intraclass correlation coefficient; LC, liquid chromatography; NMR, nuclear magnetic resonance spectroscopy; 24-HR, 24-hour dietary recall.

reflecting cereal intake, both in terms of addressing validation criteria and replication across multiple studies. Fundamental validation criteria that are most often unreported include biomarker specificity, reproducibility over time, dose response, and for some promising candidate biomarkers, correlation with habitual food intakes. Putative biomarkers and assessment of their validity per category of dietary exposures are provided below. Detailed references can be found in the narrative summaries for each dietary exposure (see [Text S1](#) in the Supporting Information online).

Biomarkers of dairy intake

There are 6 main chemical classes of biomarkers of dairy intake (ie, milk, cheese, and yogurt): (1) long-chain fatty acids and *trans*-fatty acids; (2) medium-chain fatty acids; (3) phosphatidylcholines, lysophosphatidylcholines, and cholesterol esters; (4) sugars; (5) quinolone derivatives; and (6) sphingomyelins.^{31–34} Other metabolites that do not fit into these classes include dairy additives, such as undecanoic acid.⁴¹ Caprate, a medium-chain fatty acid, has established plausibility, as it is a component in animal fat; however, because it is a component of all animal fat, and some plant and seed oils, it is nonspecific. The long-chain fatty acids pentadecanoic acid and heptadecanoic acid are synthesized by bacterial flora in ruminants; however, they are also found in meat, rendering them nonspecific to dairy. Sugars found in dairy are milk constituents, including lactose, galactose, and their metabolites, but can also be extracted from some fruits. Additionally, galactonate, a metabolite of galactose, is a product of hepatic glucose metabolism and thus can be either exogenous or endogenously derived. However, galactonate may reflect dairy intake in populations with

high intakes.⁴² Quinolone derivatives are used as antibiotics, and thus they are not specific to dairy. Blood and urine are common biospecimen sources of dairy intake biomarkers, but long-chain fatty acid dairy biomarkers are detectable in adipose tissue and erythrocytes. Dairy intake biomarkers have been analyzed from human samples using gas chromatography–MS (GC–MS), liquid chromatography–MS (LC–MS), GC coupled with a flame ionization detector (GC–FID), and NMR. Using these techniques, low to moderate correlations with dairy intake have been observed for long-chain fatty acids, such as pentadecanoic acid (15:0), myristic acid (14:0), and *trans*-palmitoleate (*trans*-16:1n-7) in both plasma and adipose tissue.^{43–52} Galactonate, sphingomyelins (SMs), and medium-chain fatty acids have low to moderate correlations with the consumption of dairy products, and lactose has a low correlation with dairy intake.^{52–55} Moderate correlations with dairy consumption have been observed for both serum cholesterol esters and 2,8-quinolinediol based on 7-day FRs.⁵⁶ Studies that reported on half-lives of these candidate dairy biomarkers were not found. Fair to good mean ICC values have been observed for heptadecanoic acid (17:0), *trans*-palmitoleic acid (*trans*-16:1n-7), and pentadecanoic acid (15:0) (ICCs ranging from 0.52 to 0.72 measured over 2 to 3 y).⁵⁷ Additionally, among fatty acid derivatives, SMs, and quinolone derivatives, good to excellent mean ICC values were also observed for *N,N,N*-trimethyl-5-aminovaleate (0.87), 3-bromo-5-chloro-2,6-dihydroxybenzoic (0.75), SMs (d17:2/16:0, d18:2/15:0) (0.65), and quinate (0.81) over 6 months.⁵² Overall, pentadecanoic acid, myristic acid (14:0), *trans*-palmitoleate, and galactonate appear to be promising candidate biomarkers of dairy intake. Other biomarkers are less specific to dairy, and therefore may be suboptimal.

Biomarkers of meat intake

Most proposed meat intake biomarkers are in the following chemical classes: (1) peptides, (2) amino acids, and (3) amino acid derivatives. They are either present in meat or formed during digestion of meat in the gut.²⁹ Examples include carnosine, acetylcarnitine, 4-hydroxyproline, 3-methylhistidine, and anserine. Trimethylamine N-oxide (TMAO), a compound that has been repeatedly associated with meat intake, is a metabolite of choline and phospholipids and a metabolite of L-carnitine. However, its use as a meat biomarker is limited in populations with fish consumption, as fish naturally contains high concentrations of TMAO.^{58,59} Other biomarkers may be specific for heated meat products, such as N-nitrosoproline formed during heating of cured meat products⁶⁰ and heterocyclic amines (MeIQx, PhIP) that are formed when amino acids react with creatinine during thermal processing of meat and meat products.⁶¹ Syringol metabolites are products of wood pyrolysis present in smoke and in smoked-meat products.⁶² Piperine and piperettine are pepper alkaloids associated with processed-meat intake (eg, sausage and salami); however, their use as a meat biomarker may be limited in populations using high amounts of pepper via other food sources than meat.⁶³

These biomarkers and biomarker precursors are measurable in meat and meat products, establishing their plausibility. Comprehensive meat-composition data are lacking in available food-composition tables and databases, making it difficult to assess specificity. The specificity of 3-methylhistidine and anserine (a dipeptide of 3-methylhistidine and alanine) has been more extensively examined than other candidate biomarkers for meat and meat product intake. Both compounds show high concentrations in chicken and low concentrations in other meats.^{64–67} However, some of these amino acids and peptides form in human tissues, which may limit their sensitivity as biomarkers. Robust correlations were observed between some biomarkers and meat intake, although some studies that identified acetylcarnitine, 4-hydroxyproline, and 3-methylhistidine as candidate biomarkers of total meat intake did not report correlation values.^{58,68} Correlations with the intake of specific meat products were also studied. 3-Methylhistidine was highly correlated with intakes of poultry/chicken and may be a biomarker of such food intakes.^{58,69,70} Syringol sulfate and piperine increased significantly across low, moderate, and high levels of habitual intake of smoked meat and sausage, respectively.^{62,63} Reproducibility over time (ICC) was fair for several compounds in urine, including for 3-methylhistidine (0.42), anserine (0.40), and acetylcarnitine (0.48).^{71,72} In blood, ICC values were poor to good for

3-methylhistidine (0.07–0.66) and poor to fair for 4-hydroxyproline (0.17–0.51), acetylcarnitine (0.34–0.55), and piperine (0.55).^{73–75} There are correlated foods that could be evaluated as potential confounders, such as smoked fish for syringol sulfate or pepper for piperine. Overall, studies have proposed a variety of biomarkers of meat intake, but many lack comprehensive validation. Biomarkers with the greatest level of validation according to the criteria include acetylcarnitine and 4-hydroxyproline for total meat intake, 3-methylhistidine and anserine for chicken intake, syringol sulfate for smoked-meat intake, and piperine for sausage intake.

Biomarkers of fish and seafood intake

Several molecules belonging to (1) furan acids, (2) fatty acids, and (3) amine oxides and their derivatives are proposed biomarkers of fish and seafood intake. 3-Carboxy-4-methyl-5-propyl-2-furanpropanoic acid (CMPF) is a metabolite formed in humans from dietary furan fatty acids, which are most abundant in fish.^{76,77} Dietary furan fatty acids have also been measured in very low concentrations in green plants, mushrooms, vegetable oils, and butter.^{76,77} However, those foods were not associated with CMPF concentrations in fasting plasma in a randomized controlled trial.⁷⁸ In cross-sectional studies in diverse free-living populations, CMPF has been associated with intakes of fish (dark, oily, and total) and shellfish, but not other foods.^{46,54,55,79} The 3 most abundant omega-3 (n-3) polyunsaturated fatty acids (PUFAs) in fish oil are eicosapentaenoic acid (EPA; *cis*-20:5n-3), docosahexaenoic acid (DHA; *cis*-22:6n-3), and docosapentaenoic acid (DPA; *cis*-22:5n-3). Accordingly, each has been associated with seafood intake and, more specifically, fish intake in human metabolomics studies. Thus, these fatty acids may both reflect fish oil and fatty fish intake. A larger number of studies have identified circulating levels of EPA and DHA, as opposed to DPA, as a candidate biomarker of fish intake.^{44,46,54,55,79,80} Additionally, 1-docosahexaenoylglycerophosphocholine (a DHA lysophosphatidylcholine and a derivative of fish oils) measured in fasted serum/plasma and nonfasted serum has been associated with total and oily fish consumption.^{46,79} Another promising candidate biomarker of fish intake is TMAO, as it is abundant in fish and seafood. However, TMAO is not seafood specific, especially in populations with low seafood intake, since, as noted above, it is associated with meat intake. The gut microbiota also generates TMAO from choline, betaine, and carnitine. However, 3 separate controlled-feeding studies found that TMAO measured in 24-hour and spot urine increased with white and fatty fish intake, or fatty fish intake alone.^{58,81,82} One of these studies

replicated the finding in a cross-sectional analysis in a subset of participants from the European Prospective Investigation into Cancer and Nutrition (EPIC) in urine and plasma.⁵⁸ No information on half-lives was available for candidate seafood intake biomarkers. High (>0.6) ICC values have been observed for serum and plasma CMPF, DHA, EPA, and TMAO (ICCs ranging from 0.55 to 0.99 measured over 4 wk to 3 y).^{52,73,83,84}

Biomarkers of vegetable intake

There are 9 classes of biomarkers of vegetable intake: (1) carotenoids, (2) tocopherols, (3) phenolic acids and derivatives, (4) flavonoids, (5) isoflavonoids, (6) retinol, (7) ascorbic acid, (8) carboxylic acids and derivatives, and (9) lipids and lipid-like molecules. While, in general, most vegetable biomarkers are associated with multiple vegetable types, there are some candidate biomarkers with more specificity. For instance, sulforaphane and S-methylcysteine are primarily found in cruciferous vegetables.⁵² Garlic is a primary source of the sulfoxide alliin and S-allylcysteine, while ergothioneine is an amino acid constituent in mushrooms. However, these compounds are not unique to these specific vegetables.^{52,54} N-acetylliin and S-allylcysteine may be more specific markers of allium vegetable intake.⁸⁵ Vegetable-related metabolites are detectable in blood, 24-hour urine, or spot urine samples using high-performance LC-MS, time-resolved fluorescent immunoassay (TR-FIA), and capillary electrophoresis–time of flight MS (CE-TOF-MS). Using these methods, blood carotenoids have shown moderate to strong correlations with vegetable intake.^{52,86–94} Of the carotenoids, only α -carotene consistently had a moderate correlation with total vegetable intake (including cooked vegetables) as well as with carrot intake.^{86–88,92–95} Additionally, carotene diol had a moderate correlation with leafy green and cruciferous vegetable intake, and β -cryptoxanthin had a moderate correlation with cucumber intake.^{52,92} Retinol had moderate correlations with both onion and leafy green vegetable intakes, and vitamin C with total vegetable, leafy green vegetable, root vegetable, and onion intakes.^{88,96} Moreover, alliin and S-allylcysteine had moderate correlations with garlic intake, and weak to moderate correlations were observed between ergothioneine and mushroom intake, N-acetylliin, and allium vegetable intake and S-methylcysteine and cruciferous vegetable intake.^{52–54,97} Studies that reported half-lives of these candidate vegetable intake biomarkers were not identified. Excellent ICCs over 6 to 12 months were observed for carotene diol (0.79–0.83), α -carotene (0.83), ergothioneine (0.86), and lutein (0.80) in 2 studies.^{52,91} Overall, the most validated dietary biomarker candidates for garlic intake are alliin and S-allylcysteine,

while ergothioneine is a potential biomarker for mushroom intake. S-Methylcysteine is a potential candidate biomarker for cruciferous vegetable intake; however, its correlation is weak, and it is also a constituent of beans.⁵² While α -carotene has a moderate correlation with total vegetable intake, it correlates with total fruit intake, rendering it a nonspecific biomarker for vegetable intake.

Biomarkers of legume (including pulses, seeds, and peanuts) and tree nut intake

Legumes (fabaceae or leguminosae) are a diverse family of flowering plants that are an important part of traditional diets worldwide. Examples of foods within the legume family include pulses (beans, lentils, and peas) and peanuts. Soy beans contain a high content of isoflavone components, including genistein and O-desmethyllangolensin (O-DMA).^{17,98} Candidate biomarkers of soy intake include blood or urine genistein and daidzein.^{99–101} Although other beans like peanuts contain low concentrations of these compounds, specificity is otherwise high.¹⁰² Both isoflavones have a relatively short plasma half-life of approximately 6 hours for genistein and 5 hours for daidzein,^{103,104} which may limit their application as soy biomarkers to populations with a frequent consumption of soy products. For example, studies in Asian populations that regularly consume soy products have demonstrated high reproducibility of soy–biomarker associations over time.^{105–108} O-DMA is a microbial metabolite formed from daidzein by the gut microbiota. However, its production depends on human gut microbial composition, limiting its use.¹⁰⁹ Furthermore, urinary excretion of O-DMA is weakly associated with soy food intake.¹¹⁰ Overall, genistein and daidzein provide robust estimates of soy intake in populations frequently consuming soy products. More limited information is available on other legumes such as peanut, different types of beans, and pulses.^{79,80,105} Guertin et al⁴⁶ and Playdon et al⁵⁵ found that the serum metabolites 4-vinylphenol sulfate and tryptophan betaine reflected peanut or nut consumption with weak correlation to habitual intake and excellent 1-year reproducibility for tryptophan betaine (ICC = 0.74). Tryptophan betaine was also associated with habitual nut intake in several other studies. 4-Vinylphenol sulfate and tryptophan betaine are both xenobiotics previously identified in roasted peanuts or legumes.⁴⁶ 2-Isopropylmalic acid, asparaginy valine, and N-carbamoyl-2-amino-2-(4-hydroxyphenyl) acetic acid were observed to increase in a dose-dependent manner with pea intake, but studies addressing correlations with habitual intakes in free-living individuals are lacking.^{111,112} Some authors have also looked at correlations

of plasma lipid metabolites, such as sphingomyelins (C24:0 and C22:0) and ceramides (C24:0), and intakes of nuts and peanuts.¹¹³ However, more studies are required to identify the plausibility and specificity of this association with nut intake. Pipecolic acid is a promising serum and urine biomarker of dry bean intake based on findings from a 4-week dietary intervention study.¹¹⁴ However, studies in larger, free-living populations are required to further establish its specificity and reproducibility.^{52,98}

Biomarkers of fruit intake

Biomarkers of fruit intake fall into 7 general classes as follows: (1) proline and derivatives, (2) flavonoids, (3) carotenoids, (4) threitol (a xylose metabolite), (5) ascorbic acid, (6) inositol isomers, and (7) dopamine sulfate. While specificity has not been determined for most of these compounds, proline betaine is specific to citrus fruit, phloretin is specific to apples, and dopamine sulfate is specific to bananas.^{115–118} Moreover, grapefruit is a source of flavanones, especially naringenin.¹¹⁹ The color of fruit depends on the type of carotenoid pigment. Notably, lycopene, which gives fruits and vegetables a red color, is a metabolite of tomatoes, and β -cryptoxanthin, which provides a yellow-orange color, is specific to certain fruits, including orange, tangerine, and papaya.^{120,121} These fruit metabolites have been measured in blood and urine, primarily using LC-MS. Moderate to strong correlations have been observed for proline derivatives with habitual intake of citrus fruits.^{46,52–54,97} Citrus fruit intake has also been observed to have a moderate to strong correlation with β -cryptoxanthin, and lycopene has a moderate to strong correlation with tomato intake among free-living individuals in population studies.^{86,88,95,122} Additionally, other carotenoids, including zeaxanthin, lutein, α -carotene, and β -carotene, measured in blood have moderate to high correlations with total fruit intake.^{87,92,123,124} Strong correlations have been observed for citrus flavonoids with citrus fruit intake, weak to moderate correlations have been reported for phloretin with apple intake, and moderate to strong correlations have been observed in both urine and plasma for dopamine sulfate and banana intake.^{54,86,95,125} Few studies have quantified the half-lives of metabolites of fruit intake. Of those assessed, flavanones and dopamine sulfate have very short half-lives (ie, <2 h).^{125–127} Fair to good ICC values have been observed for proline betaine (0.35–0.50), and good to excellent reliability was observed for various carotenoids (0.58–0.84), methyl glucopyranoside (0.62), 4-hydroxychlorothalonil (0.85), and γ -tocopherol/ β -tocopherol (0.69) over 6 months to 1 year.^{46,52,91} Using data from controlled-feeding or intervention studies, proline

betaine and flavanones showed a linear dose response to citrus intake, as did xylose with apple intake and dopamine sulfate with banana intake.^{126,128–130} Overall, proline betaine, β -cryptoxanthin, and flavanones appear to be more robust candidate biomarkers of citrus fruit intake. The strongest candidate dietary biomarker for apple intake is phloretin, whereas lycopene is a strong candidate biomarker for tomato intake. While dopamine sulfate is a banana derivative with a robust correlation with banana intake, it is also an endogenous molecule, which may limit its use.

Biomarkers of cereal food intake

Candidate biomarkers of cereal food intake include molecules associated with intakes of whole-grain wheat and rye and their bran (total alkylresorcinols [ARs], homologues C17:0–C25:0, C17:0 to C21:0 ratio, and AR metabolites 3,5-dihydroxybenzoic acid [-DHBA], 3,5-dihydroxyphenyl propanoic acid [-DHPPA], 3,5-dihydroxyphenyl pentanoic acid [-DHPPTA]), whole-grain oats (avenacoside [AVE] A and B, avenanthramides [AVAs], and their metabolites), and of sourdough fermented rye (2-hydroxy-*N*-[2-hydroxyphenyl]acetamide [HHPAA] and *N*-[2-hydroxyphenyl]acetamide [HPAA]). Moreover, metabolomics analyses have revealed several biomarker (>18 compounds) candidates including AR metabolites (DHPPA and DHPPTA), benzoxazinoid compounds or derived metabolites (2-aminophenol-sulphate, HHPAA, HMBOA, HPAA, HPPA), and microbial products of phenolic compounds (eg, hydroxybenzoic acid glucuronide, dihydroferulic acid sulfate, and enterolactone conjugates) in urine samples associated with habitual bread consumption (both whole-grain and refined-grain breads of different types) in free-living individuals.¹³¹ A set of the highest ranked candidate biomarkers were combined into a panel that predicted whole-grain bread intake with low to moderate prediction performance,¹³¹ but many of the compounds had limited or unclear specificity with whole grains from different sources vs refined grains, which may limit their usefulness until more research has been conducted. In addition, betainized compounds have been shown to increase in plasma after the consumption of whole grains (rye and wheat) under controlled conditions, and pipecolic acid betaine increased after both whole-grain wheat and rye consumption.¹³² ARs have been measured in plasma, serum, erythrocyte membranes, or adipose tissue, whereas their metabolites are measured in plasma or urine. AVAs are analyzed in plasma and AVE is analyzed in urine samples. Benzoxazinoids and their metabolites (2-aminophenol sulfate, HHPAA, and HPAA) have been analyzed in plasma and urine.^{97,133} Odd-chain ARs are mainly found in the bran of wheat and rye, but also to a

lesser extent in barley and sifted rye, with trace amounts in refined flour of wheat from contamination.^{26,134} They are not found in other food sources and are therefore specific for whole grain/bran of wheat, rye, and barley.¹³⁵ Even-numbered AR homologues are specific to quinoa intake. ARs are stable during food processing and are not degraded, as recently suggested,¹³⁶ but will form strong interactions with the matrix and need hot extraction to be released in hydrothermally produced foods such as bread.¹³⁷ Several studies compared concentrations of intact ARs in plasma with estimated intakes derived from 24-HRs, FRs, and FFQs in European and American populations. The correlation coefficients are moderate to strong depending on the population and method used to estimate intake. Plasma ARs show a linear dose–response relationship at a wide intake range.¹³⁸ A few studies have also investigated the correlation of ARs in adipose tissue biopsies with self-reported habitual intake, with similar correlations. This suggests that ARs in adipose tissue also reflect mainly short- to medium-term intake, most likely due to a rapid, dynamic turnover rate. Plasma and urine AR metabolites (3,5-DHBA, 3,5-DHPPA, 3,5-DHPPTA) in free or conjugated forms are specific to whole-grain/bran wheat and rye intake, but have also been detected after the consumption of peanuts, wort, and beer (3,5-DHBA), and after the consumption of sinapic acid and some flavonoids (3,5-DHPPA). However, the contribution of these sources is minor, and it should also be noted that some methods have wrongly identified the more common 3,4 configuration as 3,5 (3,5-DHBA and 3,5-DHPPA). Moreover, AR metabolites from spot urine and 24-hour urine show weak to moderate correlations with estimated whole-grain intake. The apparent half-life of total AR ranges from approximately 4 to 7 hours. Corresponding half-lives for AR metabolites are estimated to be 10–12 hours for 3,5-DHBA and 3,5-DHPPA and 10–16 hours for 3,5-DHBA-glycine and 3,5-DHPPTA based on plasma and urine data.^{139,140} For plasma ARs, the reproducibility over time has been shown to be fair to good over periods of 2 months to 3 years, but higher for women than for men, both for intact ARs and metabolites in plasma.^{141–143} The reproducibility of AR metabolites in plasma was similar to intact ARs, despite the longer apparent half-life.^{141–143} This may be due to unknown factors affecting the stability of AR metabolite concentrations.^{144,145}

AVAs only exist in oats and therefore have excellent specificity. They are converted by the gut microbiota into their dihydro forms,^{146,147} thus differentiating different AVA metabolotypes. AVE A and B are also highly specific to oats. Documented half-lives of AVAs range from 2.2 to 4.6 hours.¹⁴⁸ No published studies have reported the reproducibility over time of

AVAs or AVEs, plasma or urine correlations with oat intake, or dose response under controlled conditions.

Sourdough fermentation in rye generates some benzoxazinoid metabolites, but the specificity of compounds related to specific food processes remains to be elucidated. The correlation of benzoxazinoid metabolites in plasma with habitual grain intake has not been reported in population studies, but recent human intervention studies have shown large differences in their concentrations at different whole-grain (rye) intake levels, suggesting a plausible dose response.¹³³ The benzoxazinoid metabolites HHPAA and HPAA in spot urine and 24-hour urine were correlated with whole-grain rye intake in the range of $r = 0.32$ – 0.52 . One study found that a panel of biomarkers analyzed in 24-hour urine samples was associated with whole-grain rye intake.¹⁴⁹ Data on half-lives and reproducibility for benzoxazinoid and their metabolites are lacking.

In summary, ARs in plasma have been validated as biomarkers of whole-grain wheat and rye intakes and are used as such in epidemiological studies. They are highly specific, increase with increased intake in a plausible dose–response manner, and are robustly correlated with estimated whole-grain intake. What primarily limits their use is the short-half life, which makes them unsuitable as biomarkers in populations with an irregular and infrequent whole-grain intake. However, in populations with a frequent intake, such as in Scandinavian countries, ARs in plasma are feasible biomarkers of whole-grain wheat and rye intake. AR metabolites in plasma and urine have an approximately similar performance as intact ARs in plasma, despite a longer apparent half-life. Twenty-four-hour urine metabolite concentrations may be strongly correlated with estimated intake, but the feasibility of 24-hour urine collection may limit their use. AVAs and AVEs in plasma and urine appear promising as biomarkers of oat intake, but further studies to establish their pharmacokinetics and dose response under controlled intake conditions in humans as well as in observational studies are highly warranted. Similar studies are also warranted to judge the feasibility of benzoxazinoid derivatives and betainized compounds as individual biomarkers of whole-grain intake and of combinations of the most promising individual markers into biomarker panels.

Biomarkers of food component intake: sugar, alcohol, fats, and oils

Sugar. Biomarkers of sugar intake (sucrose) include (1) fructose and sucrose and (2) the isotopic signature $\delta^{13}\text{C}$, which does not belong to a specific chemical class. Sucrose originates directly from dietary sucrose, whereas fructose originates directly from dietary

fructose and is one of the monosaccharides in sucrose. With regard to $\delta^{13}\text{C}$, photosynthetic plants discriminate carbon isotopes when fixing carbon dioxide into organic molecules. This discrimination varies depending on photosynthetic pathways. C_4 plants (eg, corn, sugar cane) fix more of the heavy isotope ^{13}C than C_3 plants (most plant species). This is reflected in the $^{13}\text{C}/^{12}\text{C}$ ratio in sugars produced by these plants. The ^{13}C abundance ($^{13}\text{C}/^{12}\text{C}$ ratio; $\delta^{13}\text{C}$) is changed in human biofluids and tissues upon ingestion of sugars produced by C_4 plants. A high consumption of added sugar (eg, sucrose from sugar cane or high-fructose syrup from corn) will influence $\delta^{13}\text{C}$. Sucrose and fructose are natural constituents of many foods and food products, and the urinary concentration of these compounds will reflect both sugars naturally present in the foods as well as added sugar, and therefore not the intake of any specific food or food group. $\delta^{13}\text{C}$ reflects the intake of sugars produced by C_4 plants and has been used in the United States as a proxy for the intake of added sugars from corn and sugar cane, hence may not be applicable for other sources of sugars, like sugar beet or natural fruit sugars. Fructose and sucrose are only measured in urine, typically using GC-MS¹⁵⁰ or LC-MS,¹⁵¹ whereas $\delta^{13}\text{C}$ is measured in whole blood, red blood cells, hair, breath, and plasma by isotope-ratio MS technology. Measured $\delta^{13}\text{C}$ in hair samples and breath was significantly associated with dietary carbohydrate intake, particularly with sweetened beverages,¹⁵² and the $\delta^{13}\text{C}$ of specific amino acids, particularly alanine in serum, was moderately correlated with added-sugar intake.^{153,154} Urinary fructose and sucrose have generally shown poor to modest correlations (r range: 0.03–0.43) with habitual sugar intake estimated by FRs and somewhat weaker when using morning spot urine samples (r range: 0.20–0.30).¹⁵⁵ Despite modest correlations, the performance of urinary sucrose and fructose as biomarkers of habitual sucrose intake was comparable to urinary nitrogen as an established protein intake biomarker in a free-living Dutch adult population.¹⁵⁶ Correlations of $\delta^{13}\text{C}$ with total added-sugar and sugar-sweetened beverage intake measured by FFQs or FRs stem from studies measuring the biomarker in whole-blood samples; correlation coefficients ranged from $r=0.28$ to 0.35 depending on the exposure and dietary instrument used.¹⁵⁷ The apparent elimination half-life of sucrose is approximately 3 hours¹⁵⁸ and is 39 minutes for fructose.¹⁵⁹ The 50% turnover of $\delta^{13}\text{C}$ was reported to be 2.5 weeks in plasma and 5.9 weeks in red blood cells,¹⁶⁰ underscoring its potential to reflect long-term sugar intake. Despite the short half-life, urinary sucrose and fructose showed a modest reproducibility (ICC: 0.38–0.47) over a period of 3 years.¹⁵⁶ In summary, none of the sugar biomarkers

have been fully validated, but the currently available data suggest that sucrose and fructose in 24-hour urine, and possibly in morning spot urine, are promising biomarkers of total and extrinsic sugar intake. $\delta^{13}\text{C}$ is also a promising biomarker of habitual added sugars from C_4 plants in US populations where they are widely consumed, but this requires further validation.

Alcohol. Biomarkers of alcohol consumption are important in forensic contexts. In clinical medicine, they can verify alcohol abstinence or toxicity. Correlations between FFQ self-reported alcohol intake and ethyl glucuronide concentrations in plasma or urine are moderate to strong ($r=0.26$ –0.36 in serum; $r=0.20$ –0.60 in urine). Stronger correlations have been observed for phosphatidylethanol (PEth) measured in whole blood ($r=0.26$ –0.79). Despite a short half-life of ethyl glucuronide (ie, 2.5 h), the 6-month to 1-year reproducibility over time was reported as moderate (ICC = 0.27⁴⁶ and ICC = 0.57⁵², respectively). PEth has a longer half-life of 2–9 days depending on specific compounds evaluated, but data on its reproducibility over time are currently lacking. Other compounds may be useful biomarkers for specific alcoholic beverages. Compounds such as humulinone, isoxanthohumol, and 2,3-dihydroxy-3-methylvaleric acid have been suggested as candidate biomarkers of beer intake,^{161,162} and a combination of 7 biomarkers originating from the various raw materials used in beer production was also proposed as a biomarker of beer intake.¹⁶³ Observational studies associating self-reported beer intake with putative biomarkers of beer intake are lacking, as are data on biomarker half-lives and reproducibility over time. One study estimated the half-life of isoxanthohumol to be 20–28 hours.¹⁶⁴ Suggested biomarkers of wine intake include compounds produced during wine fermentation, small organic acids originating from the wine, and wine polyphenols and their metabolites including compounds such as resveratrol, resveratrol glucuronide and sulfates, hydroxytyrosol, tartaric acid, 2,3-butanediol, gallic acid, and gallic acid ethyl ester.^{163,165–172} The correlations between habitual self-reported intake of wine and biomarkers measured in plasma, such as 4-*O*-methylgallic acid and gallic acid ethyl ester sulfate, range from $r=0.30$ to 0.44, depending on the dietary instrument used. Correlations of urinary biomarkers with wine intake range from $r=0.22$ to 0.69. Half-lives are typically short and range between 1 and 9 hours.^{173–175} Many clinically used alcohol-exposure biomarkers have a short half-life, which limits their use in epidemiological investigations of habitual alcohol intake if consumption is sporadic.¹⁷⁶ Molecules suggested to reflect total alcohol intake include ethyl glucuronide, ethyl sulfate, 2-phenylethanol glucuronide, PEth,^{161,177,178} and more

recently, 2-hydroxy-3-methylbutyric acid.²⁰ The reproducibility has been estimated to be ICC = 0.42–0.67 for different compounds in plasma and urine.^{52,83} Most of the suggested biomarkers have been measured by LC-MS/MS and a few by GC-MS and/or NMR.

In summary, ethyl glucuronide and PEth are promising biomarkers of habitual total alcohol intake, but studies on the correlations with habitual self-reported intake in free-living populations and reproducibility over time are still lacking for PEth. Putative biomarkers of specific alcohol beverages, such as isoxanthohumol for beer intake and tartaric acid, 4-*O*-methylgallic acid, and gallic acid ethyl ester sulfate for wine intake, are promising; however, studies on the magnitude of correlation with reported intake, reproducibility, and specificity are needed. For example, 4-*O*-methylgallic acid (a human metabolite of gallic acid abundant in wine but also in tea) will not be specific enough in populations consuming tea. On the other hand, gallic acid ethyl ester, a metabolite formed in wine by esterification of ethanol (abundant in wine and absent in tea) with gallic acid, is more specific to wine intake.

Fats and oils. Biomarkers of fat and oil intake include several chemical classes: (1) unsaturated fatty acids, (2) saturated fatty acids, and (3) amino acid derivatives. Pentadecanoic acid or pentadecylic acid (15:0), heptadecanoic acid or margaric acid (17:0), and myristic acid (14:0) are long-chain fatty acids typically associated with butter consumption.^{43,46,51,56,79,179,180} Palmitelaidic acid or *trans*-16:1n-7 is also a long-chain fatty acid typically found in butter and margarine. Very-long-chain fatty acids include EPA (*cis*-20:5n-3), DHA (*cis*-22:6n-3), and DPA (*cis*-22:5n-3) and other n-3 PUFAs. EPA and DHA are the 2 most abundant n-3 fatty acids in fish oil and in marine mammal fat,¹⁸¹ and DHA is the third most abundant long-chain n-3 fatty acid in fish oil.¹⁸¹ These fatty acids are all detectable in blood. Other biomarkers associated with fat/oil intake, such as creatine, N-acetylglutamine, and N-acetyltyrosine, have been measured in overnight urine samples and have been associated with different types of fats and oils, including butter, margarine, meat fat, mayonnaise, salad dressing, oil used for cooking, and shortening.⁵⁵ Pentadecanoic acid (15:0) and heptadecanoic acid (17:0) are synthesized by bacterial flora in ruminants and are not produced in humans, whereas very-long-chain fatty acids (eg, EPA, DHA, DPA, and other n-3 PUFAs) are found in fish oil. Correlations of fat/oil intake with EPA, DHA, and DPA have been observed in serum and plasma (fasted and nonfasted).

Correlations of fat and oil biomarkers with habitual dietary intake ranged up to 0.19 for 17:1, 0.40 for heptadecanoic acid (17:0), 0.47 for pentadecanoic acid (15:0),

and 0.22 for *trans*-palmitoleic acid (*trans*-16:1n-7). Reproducibility (ICC) over time was moderate for 17:0 (0.50), 16:1 (0.60), and 15:0 (0.70). Most of these biomarkers are constituents of related foods, like fish and fish products for the very-long-chain fatty acids and other dairy products for 15:0 and 17:0.

Overall, a variety of blood and urine biomarkers for the intake of fats and oils have been proposed, although few have been fully validated. Some of the biomarkers of fats and oils also reflect specific food intakes such as marine foods (long-chain n-3 PUFAs) and dairy foods (C15:0 and C17:0). Urine creatine, N-acetylglutamine, and N-acetyltyrosine are promising fat and oil biomarkers. Blood DHA (*cis*-22:6n-3), DPA (*cis*-22:5n-3), EPA (*cis*-20:5n-3), n-3 PUFAs, margaric acid (17:0), methyl palmitic acid (C17H34O2), and pentadecylic acid (15:0) are also promising.

Biomarkers of tea intake

Candidate biomarkers of tea intake include the following: (1) gallic acid and its derivatives, (2) catechins and catechin metabolites, (3) carboxylic acid and its derivatives, and (4) flavonoids. With regard to the plausibility of these biomarkers, catechins, gallic acid, and flavonols are phenolic compounds found in tea leaves, with the amount depending on the variety of tea.¹⁸² The amino acid theanine is also a constituent in tea.⁵⁴ Catechins, such as epigallocatechin and epicatechin, are particularly abundant in tea. Catechins are also present in other foods, however, such as fruits, chocolate, some vegetables, and nuts. Similarly, chocolate and wine contain gallic acid, which makes the biomarker specific only for some populations depending on intakes of these foods.^{183,184} Theanine, while primarily found in tea, is also found in mushrooms. Tea-associated metabolites have been detected in blood and 24-hour urine by LC-MS and CE-TOF-MS, and in spot urine with analysis by LC-MS. Moderate to high correlations with habitual tea intake have been observed for urine gallic acid metabolites 4-*O*-methylgallic acid and methylgallic acid sulfate.^{184,185} Additionally, theanine measured in blood had moderate correlations with tea intake, and the correlations tended to be higher in caffeinated or nonherbal teas.^{52–54} Although epigallocatechin, epicatechin, and other catechin metabolites have shown dose-response relationships to tea intake in interventional studies, they show low to moderate correlations with habitual tea intake measured by FFQs in population studies.^{95,184} Kaempferol was the only flavonoid showing a moderate correlation with habitual tea intake.^{86,95} Moderate ICC values were observed for 3-methoxycatechol sulfate and theanine.⁵² Overall, 4-*O*-methylgallic acid, methylgallic acid sulfate, and theanine appear to

be the most promising candidate biomarkers for tea intake, despite some limitations in their specificity.

Biomarkers of coffee intake

Coffee intake biomarkers derive from several chemical classes: (1) caffeine and its metabolites, (2) phenolic acids, (3) organic acids, (3) niacin derivatives, and (4) roasting compounds, among others. Coffee beans and coffee brews contain a number of characteristic compounds, including caffeine, 5-caffeoylquinic acid (chlorogenic acid), feruloylquinic acid, and trigonelline, which is derived from the metabolism of niacin (vitamin B₃). These compounds are eventually transformed by the gut microbiota or human tissues into a number of metabolites such as theophylline and theobromine (from caffeine), caffeic acid, dihydrocaffeic acid and quinic acid (from chlorogenic acid), or nicotinic acid (from trigonelline). In addition to these compounds naturally present in coffee beans, other compounds like diketopiperazines and *N*-(2-furoyl)glycine are formed during roasting of beans and show increased levels in highly roasted coffee beans.¹⁸⁶ All of these compounds can be absorbed in the gut and are found in blood and urine after coffee consumption. Some of these compounds can also be found in other dietary sources (eg, caffeine in tea). However, for most compounds, their concentrations in other foods or beverages are much lower in comparison with coffee, resulting in good specificity as biomarkers of coffee intake. Most biomarker studies have focused on biomarkers of intake of generic coffee (ie, any coffee type and processing method). However, combinations of biomarkers could be used to study intakes of particular types of coffee, such as the caffeine and trigonelline ratio for caffeinated over decaffeinated coffee.^{187,188}

Coffee biomarkers are detectable in both blood (plasma or serum) and urine. Caffeine and its metabolites have low to moderate correlations with habitual coffee intake. Trigonelline and quinic acid have been most strongly correlated with coffee intake (*r* values up to 0.61 and 0.77, respectively).¹⁸⁷ Among European populations, biomarkers showing the highest correlations with coffee consumption were found to vary, possibly reflecting the different types of coffee brews consumed in each country.¹⁸⁸ Most coffee biomarkers have short half-lives (maximum of 5 h). However, trigonelline and quinic acid have high mean ICC values over time (0.66 and 0.81 over 6–12 mo, respectively), likely due to the frequency of coffee consumption.^{52,189}

Overall, trigonelline and quinic acid appear to be better qualified biomarkers of coffee intake than other candidates. Combinations of these biomarkers with compounds such as caffeine, diketopiperazines, and *N*-(2-furoyl)glycine should be tested in future studies as they may provide information on the type of coffee consumed.

Summary of the extent of candidate dietary biomarker validation

The extent of validation of selected dietary biomarkers that appear to be the most promising candidates based on the validation criteria is summarized in Table 2. All top candidate biomarkers have been studied in either serum or plasma and, in some cases, in urine. The meat biomarker acetylcarnitine and the tea biomarker 4-*O*-methylgallic acid were studied in urine samples only. The top legume, fish/seafood, whole-grain, and coffee biomarkers are specific to those food groups, whereas others may be associated with 1 or more other dietary exposures. The reproducibility over time has been investigated for most of the top dietary biomarkers, being moderate for the top vegetable, legume, fish/seafood, and coffee biomarkers. For other biomarkers, the reproducibility over time ranged from weak to moderate. As expected, the magnitude of correlation of the top dietary biomarkers with short-term food intake (ie, measured by 24-HRs or FRs) tended to be stronger than the correlation with long-term food intake (ie, measured by FFQs). The strongest correlations with habitual food intake per biomarker ranged from 0.28 (ergothioneine) to 0.62 (genistein). None of the top dietary biomarker candidates have recognized nondietary determinants, such as major confounding factors or effect modifiers, except for ARs, which were shown to have higher and more variable plasma concentrations for men than for women, and stronger ICCs over time for women than for men.¹⁴³ All top dietary biomarkers can be measured by both LC-MS and GC-MS; acetylcarnitine and 3-methylhistidine can also be captured by NMR. Finally, few of the top biomarkers have been evaluated for dose response with the food source. Steps to fully validate this panel of dietary biomarkers include further studies on reproducibility over time and dose-response feeding studies.

CONCLUSION

Many biomarkers of the foods and food components outlined in the current review have been suggested over the years, but very few have been fully validated. The most promising biomarkers of each food category assessed are listed in Table 3, which notes critical gaps needed to be addressed to be considered validated according to criteria outlined by the FoodBALL consortium, which were modified to consider criteria specific for epidemiological studies.²⁵

Table 2 Candidate dietary biomarkers with the highest level of validation

Food group and most validated candidate biomarkers	Biospecimen	Specificity	Parent compound	Reproducibility over time (ICC)	Correlation (<i>r</i>) with intake measured by 24-HR or FR	Correlation (<i>r</i>) with intake measured by FFQ	Nondietary determinants	Analytical platform	Dose response
Vegetable									
Ergothioneine (mushroom)	Serum/plasma	No (organ meats, beans)	No	0.86	0.26	0.19–0.28	No	LC/GC	NA
α -Carotene (total vegetable)	Serum/plasma	No (fruit)	Yes	0.83	0.04–0.52	0.17–0.50	No	LC/GC	NA
Fruit									
Proline betaine	Serum/plasma	No (some cheese, alfalfa, artichoke, seafood)	No	0.35–0.50	0.25–0.77	0.15–0.55	No	LC/GC	NA
Proline betaine	Urine	No (some cheese, alfalfa, artichoke, seafood)	No	NA	0.48–0.80	0.43–0.50	No	LC/GC	Yes
β -Cryptoxanthin	Serum/plasma	No (total fruit)	Yes	0.50–0.77	0.04–0.49	0.11–0.57	No	LC/GC	NA
Legumes									
Genistein	Serum/plasma	Yes	Yes	0.22–0.93	0.51	0.33–0.62	No	LC/GC	NA
Genistein	Urine	Yes	Yes	0.33	0.53	NA	No	LC/GC	NA
Daidzein	Serum/plasma	Yes	Yes	0.11–0.92	0.49	0.09–0.49	No	LC/GC	NA
Daidzein	Urine	Yes	Yes	0.17	0.44	NA	No	LC/GC	NA
Dairy									
Pentadecanoic acid	Serum/plasma	No (meat)	Yes	0.72	0.17–0.54	0.03–0.36	No	LC/GC	NA
Pentadecanoic acid	RBCs	No (meat)	Yes	NA	NA	0.06–0.30	No	LC/GC	NA
Pentadecanoic acid	Adipose tissue	No (meat)	Yes	NA	0.16–0.72	0.09–0.39	No	LC/GC	NA
<i>trans</i> -9-Hexadecenoic acid	Serum/plasma	No (partially hydrogenated vegetable oil, red meat)	Yes	0.57	NA	0.06–0.39	No	LC/GC	NA
<i>trans</i> -9-Hexadecenoic acid	RBCs	No (partially hydrogenated vegetable oil, red meat)	Yes	NA	NA	0.07–0.32	No	LC/GC	NA
Meat									
Acetylcarnitine	Urine	NA	Yes	0.48	NA	0.26–0.32	No	LC/GC/NMR	Yes
4-Hydroxyproline	Serum/plasma	No (fish)	Yes	NA	NA	0.075–0.36	No	LC/GC	NA
Chicken									
3-Methylhistidine	Serum/plasma	NA	NA	0.07–0.66	NA	0.40–0.54	No	LC/GC/NMR	NA
Fish/shellfish									
CMPF	Serum/plasma	Yes	No	0.33–0.99	0.47	0.24–0.47	No	LC/GC	NA
CMPF	Urine	Yes	No	NA	NA	0.26–0.27	No	LC/GC	NA
DHA	Serum/plasma	Yes	Yes	0.55–0.95	NA	0.26–0.37	No	LC/GC	NA
DHA	Adipose tissue	Yes	Yes	NA	NA	0.15	No	Other	NA
Whole-grain wheat and rye (tabulated information for biomarker candidates of other cereals are missing)									
Alkylresorcinols	Serum/plasma, adipose tissue	Yes	Yes	0.42–0.64	0.40–0.55	0.17–0.39	Sex	LC/GC	Yes
DHBA	Plasma/urine	Yes	No	0.17–0.55	0.35–0.57	0.10–0.43	No	LC/GC	Yes
DHPPA	Plasma/urine	Yes	No	0.22–0.45	0.26–0.57	0.23–0.46	No	LC/GC	Yes
DHPPTA	Plasma/urine	Yes	No	0.63	0.17–0.52	NA	No	LC/GC	NA

(continued)

Table 2 Continued

Food group and most validated candidate biomarkers	Biospecimen	Specificity	Parent compound	Reproducibility over time (ICC)	Correlation (r) with intake measured by 24-HR or FR	Correlation (r) with intake measured by FFQ	Nondietary determinants	Analytical platform	Dose response
Tea									
4-O-Methylgallic acid	Urine	No (wine, nuts, some berries)	No	NA	0.14–0.62	0.41–0.50	No	LC/GC	NA
Theanine	Serum/plasma	No (mushroom)	Yes	0.60	0.28–0.51	0.23–0.50	No	LC/GC	NA
Coffee									
Trigonelline	Serum/plasma	Yes	Yes	0.66–0.84	NA	0.12–0.66	No	LC/GC	NA
Quinic acid	Serum/plasma	Yes	No	0.81	0.74	0.16–0.77	No	LC/GC	NA
Fats and oils									
EPA (<i>cis</i> -20:5n-3)	Serum/plasma	Yes	Yes	NA	NA	0.29–0.44	No	LC/GC	NA
DPA (<i>cis</i> -22:5n-3)	Serum/plasma	Yes	Yes	NA	NA	0.24–0.38	No	LC/GC	NA
DHA (<i>cis</i> -22:6n-3)	Serum/plasma	Yes	Yes	NA	NA	0.25–0.36	No	LC/GC	NA
Alcohol									
PEth	Whole blood	Yes (total alcohol)	Yes	NA	NA	0.26–0.79	No	LC	NA
Ethyl glucuronide	Urine	Yes (total alcohol)	No	0.27–0.57	NA	0.20–0.60	No	LC	NA
Sugar									
Sucrose	Urine	Yes (sucrose)	Yes	0.38–0.47	0.20–0.30	NA	No	LC/GC	Yes
$\delta^{13}\text{C}$	Blood	Yes (C4 plant-derived sugars)	Yes	NA	NA	0.28–0.35	No	GC/LC	NA

Abbreviations: CMPF, 3-carboxy-4-methyl-5-propyl-2-furanpropanoate; DHA, docosahexaenoic acid (22:6n-3); DHBA, dihydroxybenzoic acid; DHPPA, dihydroxyphenyl propanoic acid; DHPPTA, dihydroxyphenyl pentanoic acid; DPA, docosapentaenoic acid; EPA, eicosapentaenoic acid; FR, food record; GC, gas chromatography; ICC, intraclass correlation coefficient; LC, liquid chromatography; NA, not available; NMR, nuclear magnetic resonance spectroscopy, PEth, phosphatidylethanol; RBC, red blood cell; 24-HR, 24-hour dietary recall.

Table 3 Dietary biomarkers and validation criteria yet to be addressed

Dietary exposure	Dietary biomarker(s)	Information lacking for the biomarkers
Dairy	Pentadecanoic acid (15:0), myristic acid (14:0), <i>trans</i> -palmitoleate (<i>trans</i> -16:1n-7), and galactonate	Additional studies are needed to confirm myristic acid as a positive biomarker for dairy intake, and data on <i>trans</i> -palmitoleate's correlation with habitual intake is lacking. Dose response has also not been evaluated for any of these biomarkers.
Meat	Acetylcarnitine and 4-hydroxyproline for total meat intake, 3-methylhistidine and anserine for chicken intake, syringol sulfate for smoked meat, and piperine for sausage	Limited correlation values with habitual meat intake have been published so far and replication in different populations is needed. More data on specificity are also needed.
Fish and seafood	CMPF for lean and total fish, EPA and DHA for fatty fish	Reproducibility over time is lacking for fish intake biomarkers and also correlation of intake with different types of fish across different populations for CMPF.
Vegetables	Alliin and S-allylcysteine in blood for garlic; N-acetyllalliin in blood for allium vegetables and ergothioneine in blood for mushrooms; α -carotene in blood for total vegetable intake	Dose response has not been assessed for any of these 5 biomarkers. Reproducibility data are also lacking for alliin.
Legumes	Genistein and daidzein in blood and urine for soy and soy product intake in frequent consumers	More studies needed on pipercolic acid as a candidate biomarker for dry bean intake.
Fruits	Proline betaine in blood and urine as well as flavanones in urine for citrus fruit intake and β -cryptoxanthin in blood for tropical fruits; phloretin in urine for apples, lycopene in blood for tomato, and dopamine sulfate in blood, and urine for bananas; inositol in blood and urine is a promising biomarker for total fruit intake	Dose response has not been demonstrated for phloretin, proline betaine in blood, or carotenoids. No data are available for reproducibility of proline betaine in urine.
Cereals	Alkylresorcinols and their main metabolites DHPA, DHPPTA in plasma and urine for whole-grain wheat and rye intake, and AVAs and AVEs for oat intake	Estimation of correlations between habitual whole-grain intake and DHPPTA, AVAs, and AVEs in plasma and urine as well as estimation of their reproducibility are lacking. Moreover, estimations of half-lives of AVEs are also lacking.
Sugar	Fructose and sucrose in 24-h urine collections as well as $\delta^{13}\text{C}$ in whole blood	Evaluation of the correlations of fructose and sucrose in morning or spot urine samples with self-reported intake is warranted. Estimations of reproducibility are scarce for all 3 candidate biomarkers in all different matrices.
Alcohol	Ethyl glucuronide and PEth for total alcohol, isoxanthohumol for beer intake, and tartaric acid and gallic acid ethyl ester sulfate for wine	Studies on the correlations between ethyl glucuronide and PEth with self-reported intake and estimations of reproducibility are currently lacking and are warranted. Putative biomarkers of specific alcohol beverages such as isoxanthohumol for beer intake and tartaric acid and gallic acid ethyl ester sulfate for wine intake require further evaluation with regard to estimation of their sensitivity and specificity in free-living subjects. Their reproducibility also needs to be assessed.
Tea	4-O-Methylgallic acid and methylgallic acid sulfate in urine, and theanine in blood	Dose response has not been assessed for any of these 3 promising tea biomarkers. Only theanine has data on reproducibility. For gallic acid metabolites, possible confounding with wine intake needs to be assessed.
Coffee	Trigonelline and quinic acid in blood and urine.	Combinations with other coffee biomarkers may provide details on the type of coffee beverage consumed.
Fats and oils	Blood/plasma fatty acids, particularly long-chain polyunsaturated fatty acids, are relatively good biomarkers for the consumption of plant-based oils and fats. Very-long-chain fatty acids are promising biomarkers for oils and fats derived from seafood (including fish and marine mammals)	Dose response has not been assessed for any of these biomarkers.

Abbreviations: AVA, avenanthramide; AVE, avenacoside; CMPF, 3-carboxy-4-methyl-5-propyl-2-furanpropanoate; DHA, docosahexaenoic acid (22:6n-3); DHPA, dihydroxyphenyl propanoic acid; DHPPTA, dihydroxyphenyl pentanoic acid; EPA, eicosapentaenoic acid; PEth, phosphatidylethanol.

Many of the most promising dietary biomarkers discussed have short to medium half-lives, ranging from 4 hours to several days. Despite this, some showed modest to good reproducibility. This is likely due to frequent the intake of the reference foods, which compensates for the short half-life in providing a stable concentration in biospecimens like blood or urine. Averaging biomarker concentrations from repeated biospecimen sampling could attenuate fluctuations in a dietary biomarker concentration that results from having a short half-life and modest reproducibility. The development of simple sampling techniques that can be performed at home, such as dried blood and urine spots, would enhance feasibility to accomplish repeated sampling on a large scale.¹⁹⁰ Moreover, the small sample volumes collected with these techniques may require the development of new analytical methods. The development of novel quantitative methods to measure more comprehensive biomarker panels will also be required. Such methods could provide more efficient use of sample volumes and be more cost-effective.

Only a few candidate dietary intake biomarkers have long half-lives and could represent long-term intake in the absence of regular consumption. Biomarkers with modest to excellent reproducibility over time (measured as ICCs) were identified, suggesting a single measurement could be used to reflect long-term intake (Tables 2 and 3). Moreover, analysis of biomarker candidates in other matrices than blood compartments may reflect more long-term intake. For example, analysis of odd-chain fatty acids and carotenoids in adipose tissue biopsy samples has been shown to correlate well with long-term dairy intake³¹ and fruit intake,¹⁹¹ respectively. Other biospecimens such as hair may reflect long-term food intake as molecules tend to be retained in hair, but the development of dietary biomarkers from hair samples has rarely been attempted¹⁹² and may be limited for use in certain populations (eg, non-bald, no use of hair products containing chemicals like dye, etc).¹⁹³ The formation of adducts with blood proteins or DNA may be another option for long-term reflection of dietary intake, since the half-life of DNA adducts are generally longer than for circulating compounds.¹⁹⁴ However, the number of food-specific adducts is limited and more exploratory studies would be needed for biomarker discovery. Recent studies have shown that microRNAs from plant-based foods are absorbed in humans to some degree and are detectable in blood samples. They have been discussed as food intake biomarkers, but it is unlikely that they will reflect long-term intake.¹⁹⁵

Although several promising dietary biomarkers have been characterized, there is still substantial work needed both to discover new biomarkers for critical

foods, such as sugar-sweetened beverages, and to provide complete validation. Dietary intervention studies are needed to address that lack of dose-response data available. Many biomarker candidates are metabolites formed in the body from parent food compounds, and characterization of the factors influencing their formation is needed. Moreover, some biomarkers may reflect several food groups, such as biomarkers of fruits may also reflect the intake of fruit juices. Estimates of biomarker reproducibility in free-living populations are often missing. The field would benefit from characterization of biomarker variability and factors affecting such variability. This information will be essential to evaluate the size of populations and the number of repeated biospecimen collections needed to study their associations with health and disease outcomes in cohort studies.⁷³ In some cases, fundamental data on biomarker correlation with self-reported dietary intake are also missing. The most comprehensively evaluated biomarkers include proline betaine (a biomarker of citrus intake) and ARs and their metabolites (biomarkers of whole-grain wheat and rye intake), and this is reflected by their more routine use in nutritional epidemiology. Yet, major gaps exist in the validation of other promising dietary biomarker candidates.

In most cases, dietary exposures have been reflected by single biomarkers. Although it may be practical to analyze fewer biomarkers, single molecules may lack specificity for the exposure of interest. Biomarker panels that jointly reflect individual foods, food groups, or dietary patterns have the potential to mitigate this issue.¹⁹⁶ Combinations of diet-derived molecules with varying proportions in different food sources could also increase biomarker specificity. Comparing several biomarkers simultaneously could shed light on the specificity of multiple dietary biomarker profiles. Several blood metabolite signatures have been associated with adherence to specific dietary patterns.¹⁹⁷ However, it is yet unclear to what extent such signatures reflect the food components of the dietary pattern per se, or if they are derived from interactions with other environmental exposures, individual or lifestyle factors, human endogenous metabolism, or gut microbiota. The field could benefit from a framework for the validation of biomarkers of dietary patterns and their interpretation.¹⁹⁷

Although dietary biomarkers are promising to objectively assess dietary intake, methodological limitations, such as potential nondietary determinants; poor reproducibility due to random error associated with episodic consumption; sample instability due to collection, processing, or storage method; and analytical drift of the response of the mass spectrometer along the analysis of large series of samples that induces measurement error. Such limitations thus render the biomarkers

suitable as a complement to traditional self-reported dietary assessment, rather than as an alternative. Methods to combine biomarker measurements with traditional dietary assessments can improve the precision in the ranking of intake of specific foods in observational studies, and can be used to calibrate self-reported data.¹⁹⁶ To date, dietary data calibration has leveraged doubly-labeled water and urinary nitrogen as recovery biomarkers for energy and protein intakes, respectively. However, non-recovery biomarkers—namely, concentration biomarkers such as carotenoids, tocopherols, folate, vitamin B₁₂, and phospholipid fatty acids—have more recently been shown to be useful to correct for systematic measurement error in self-reported nutrient intake when assessing diet and disease associations.^{23,198} This has opened the door for biomarkers beyond recovery biomarkers; thus, concentration dietary biomarkers described in the present review have the potential to correct measurement errors by calibration and to improve subject ranking of estimated food intake. For example, proline betaine recently corrected measurement errors in self-reported dietary data using a calibration approach,¹⁹⁸ and plasma ARs were successfully used in combination with whole-grain intake data from FFQs to improve precision in the ranking of whole-grain intake in relation to colorectal cancer incidence.¹⁹⁹ It has been posited that biomarker measures in approximately 30% of large-study populations could be adequate to generate calibration equations.

Analyses of single biomarkers as well as panels have often been conducted with a wide variety of analytical methods, which makes interpretations more difficult due to differences in results. There is a need for comprehensive, simple, and robust assays for analysis of dietary biomarkers that can be widely adopted.

In summary, efforts have been made to discover, and to a lesser extent, validate dietary biomarkers during the last 10 years. Separate comprehensive review articles on candidate biomarkers of specific food intakes have been published recently by the FoodBALL consortium and by other authors, but to our knowledge, this review provides the first comprehensive assessment of the emerged biomarker candidates according to established validation criteria adapted for epidemiological studies. This review identified specific gaps related to the validation of specific biomarkers as well as general developments needed to take the application of dietary biomarkers further in the field of nutritional epidemiology. Future studies that emphasize the validation of individual biomarkers, biomarker panels, and the development of analytical methods that capture many dietary biomarkers in a single analysis are warranted. Moreover, evaluation of their use together with other

dietary assessment methods should also be further studied. There is also a need to better understand the impact of fasting status and timing of sampling for the validity and reproducibility of biomarker measurements; this is as yet unknown for most biomarkers, although it may to some degree be predictable from kinetics data. Another area for future research is to find specific biomarkers of food preparation and processing, since they may have health implications. Finally, data-driven or predefined panels of biomarkers that reflect specific dietary patterns and whole diets would be useful for future epidemiological investigations and there are promising developments in this area under way.¹⁹⁷ With further developments, the field of nutritional epidemiology is therefore poised to benefit dramatically from improved dietary intake assessment, which will serve to strengthen the validity of studies on diet, health, and disease.

Acknowledgments

Author contributions. R.L., E.L., I.H., M.C.P., and A.S. designed the research (project conception, development of the overall research plan, and study oversight). All authors collected and analyzed the data. All authors wrote the manuscript. R.L., E.L., I.H., M.C.P., and A.S. had primary responsibility for final content. All the authors read and approved the final manuscript.

Funding. National Cancer Institute (5R00CA218694-03, to M.C.P.); the Huntsman Cancer Institute Cancer Center (P30CA040214, to M.C.P.); Swedish Research Council (2019-01264, to S.N.); FORMAS (2018-01044, to R.L.) and FORMAS (2019-02201, to S.N.) under the umbrella of the European Joint Programming Initiative “A Healthy Diet for a Healthy Life” (JPI HDHL) and of the ERA-NET Cofund HDHL-INeSTInal MICrobiome (INTIMIC) (GA no. 727565 of the EU Horizon 2020 Research and Innovation Programme; to R.L. and S.N.); and the National Cancer Institute (NCI) Intramural Research Program (to E.L.).

Declaration of interest. The authors have no relevant interests to declare.

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 or the National Cancer Institute.

Supporting Information

The following Supporting Information is available through the online version of this article at the publisher's website.

Table S1 A summary of candidate dietary biomarkers identified per food category, assessment of them according to validation criteria, and their references.

Text S1 Standardized summary sheets of the most promising candidate biomarkers per food category and appraisal of them according to validation criteria along with key references.

REFERENCES

- GBD Collaborators. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2019;393:1958–1972. doi:10.1016/S0140-6736(19)30041-8
- Satija A, Yu E, Willett WC, et al. Understanding nutritional epidemiology and its role in policy. *Adv Nutr*. 2015;6:5–18. doi:10.3945/an.114.007492
- Dao MC, Subar AF, Warthon-Medina M, et al. Dietary assessment toolkits: an overview. *Public Health Nutr*. 2019;22:404–418. doi:10.1017/S1368880018002951
- Bingham SA. Biomarkers in nutritional epidemiology. *Public Health Nutr*. 2002;5:821–827. doi:10.1079/phn2002368
- Maruvada P, Lampe JW, Wishart DS, et al. Perspective: dietary biomarkers of intake and exposure—exploration with omics approaches. *Adv Nutr*. 2020;11:200–215. doi:10.1093/advances/nmz075
- Kaaks R, Riboli E, Sinha R. Biochemical markers of dietary intake. *IARC Sci Publ*. 1997;142:103–126.
- Jenab M, Slimani N, Bictash M, et al. Biomarkers in nutritional epidemiology: applications, needs and new horizons. *Hum Genet*. 2009;125:507–525. doi:10.1007/s00439-009-0662-5
- Gao Q, Praticò G, Scalbert A, et al. A scheme for a flexible classification of dietary and health biomarkers. *Genes Nutr*. 2017;12:34. doi:10.1186/s12263-017-0587-x
- Brennan L, Hu FB. Metabolomics-based dietary biomarkers in nutritional epidemiology—current status and future opportunities. *Mol Nutr Food Res*. 2019;63:e1701064. doi:10.1002/mnfr.201701064
- Scalbert A, Brennan L, Manach C, et al. The food metabolome: a window over dietary exposure. *Am J Clin Nutr*. 2014;99:1286–1308. doi:10.3945/ajcn.113.076133
- Knaze V, Rothwell JA, Zamora-Ros R, et al. A new food-composition database for 437 polyphenols in 19,899 raw and prepared foods used to estimate polyphenol intakes in adults from 10 European countries. *Am J Clin Nutr*. 2018;108:517–524. doi:10.1093/ajcn/nqy098
- Neveu V, Nicolas G, Salek RM, et al. Exposome-Explorer 2.0: an update incorporating candidate dietary biomarkers and dietary associations with cancer risk. *Nucleic Acids Res*. 2020;48:d908–d912. doi:10.1093/nar/gkz1009
- Wishart DS, Feunang YD, Marcu A, et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res*. 2018;46:d608–d617. doi:10.1093/nar/gkx1089
- Brouwer-Brolsma EM, Brennan L, Dreven CA, et al. Combining traditional dietary assessment methods with novel metabolomics techniques: present efforts by the Food Biomarker Alliance. *Proc Nutr Soc*. 2017;76:619–627. doi:10.1017/s0029665117003949
- Jawhara M, Sørensen SB, Heitmann BL, et al. Biomarkers of whole-grain and cereal-fiber intake in human studies: a systematic review of the available evidence and perspectives. *Nutrients*. 2019;11:2994. doi:10.3390/nu11122994
- Ulaszewska MM, Weinert CH, Trimigno A, et al. Nutrimetabolomics: an integrative action for metabolomic analyses in human nutritional studies. *Mol Nutr Food Res*. 2019;63:e1800384. doi:10.1002/mnfr.201800384
- Rafiq T, Azab SM, Teo KK, et al. Nutritional metabolomics and the classification of dietary biomarker candidates: a critical review. *Adv Nutr*. 2021;12:2333–2357. doi:10.1093/advances/nmab054
- Biskup I, Kyrø C, Marklund M, et al. Plasma alkylresorcinols, biomarkers of whole-grain wheat and rye intake, and risk of type 2 diabetes in Scandinavian men and women. *Am J Clin Nutr*. 2016;104:88–96. doi:10.3945/ajcn.116.133496
- Shi L, Brunius C, Johansson I, et al. Plasma metabolite biomarkers of boiled and filtered coffee intake and their association with type 2 diabetes risk. *J Intern Med*. 2020;287:405–421. doi:10.1111/joim.13009
- Lofffield E, Stepien M, Viallon V, et al. Novel biomarkers of habitual alcohol intake and associations with risk of pancreatic and liver cancers and liver disease mortality. *J Natl Cancer Inst*. 2021;113:1542–1550. doi:10.1093/jnci/djab078
- Lofffield E, Rothwell JA, Sinha R, et al. Prospective investigation of serum metabolites, coffee drinking, liver cancer incidence, and liver disease mortality. *J Natl Cancer Inst*. 2020;112:286–294. doi:10.1093/jnci/djz122
- Dragsted LO, Gao Q, Praticò G, et al. Dietary and health biomarkers—time for an update. *Genes Nutr*. 2017;12:24–24. doi:10.1186/s12263-017-0578-y
- Lampe JW, Huang Y, Neuhauser ML, et al. Dietary biomarker evaluation in a controlled feeding study in women from the Women's Health Initiative cohort. *Am J Clin Nutr*. 2017;105:466–475. doi:10.3945/ajcn.116.144840
- Landberg R, Hanhineva K. Biomarkers of a healthy Nordic diet—from dietary exposure biomarkers to microbiota signatures in the metabolome. *Nutrients*. 2019;12:27. doi:10.3390/nu12010027
- Dragsted LO, Gao Q, Scalbert A, et al. Validation of biomarkers of food intake—critical assessment of candidate biomarkers. *Genes Nutr*. 2018;13:14. doi:10.1186/s12263-018-0603-9
- Landberg R, Hanhineva K, Tuohy K, et al. Biomarkers of cereal food intake. *Genes Nutr*. 2019;14:28. doi:10.1186/s12263-019-0651-9
- Clinton SK, Giovannucci EL, Hursting SD. The World Cancer Research Fund/American Institute for Cancer Research Third Expert Report on Diet, Nutrition, Physical Activity, and Cancer: impact and future directions. *J Nutr*. 2020;150:663–671. doi:10.1093/jn/nxz268
- Dragsted LO. Biomarkers of meat intake and the application of nutrigenomics. *Meat Sci*. 2010;84:301–307. doi:10.1016/j.meatsci.2009.08.028
- Cuparencu C, Praticò G, Hemeryck LY, et al. Biomarkers of meat and seafood intake: an extensive literature review. *Genes Nutr*. 2019;14:35. doi:10.1186/s12263-019-0656-4
- Andersen LF, Solvoll K, Dreven CA. Very-long-chain n-3 fatty acids as biomarkers for intake of fish and n-3 fatty acid concentrates. *Am J Clin Nutr*. 1996;64:305–311. doi:10.1093/ajcn/64.3.305
- Wolk A, Vessby B, Ljung H, et al. Evaluation of a biological marker of dairy fat intake. *Am J Clin Nutr*. 1998;68:291–295. doi:10.1093/ajcn/68.2.291
- Münger LH, Trimigno A, Picone G, 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*. 2017;16:3321–3335. doi:10.1021/acs.jproteome.7b00319
- Rouge P, Cornu A, Biesse-Martin AS, et al. Identification of quinoline, carboline and glycinamide compounds in cow milk using HRMS and NMR. *Food Chem*. 2013;141:1888–1894. doi:10.1016/j.foodchem.2013.04.072
- Rodríguez-Gómez R, García-Córcoles MT, Cipa M, et al. Determination of quinoline residues in raw cow milk. Application of polar stir-bars and ultra-high performance liquid chromatography-tandem mass spectrometry. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess*. 2018;35:1127–1138. doi:10.1080/19440049.2018.1430382
- Rothwell JA, Madrid-Gambin F, García-Aloy M, et al. Biomarkers of intake for coffee, tea, and sweetened beverages. *Genes Nutr*. 2018;13:15. doi:10.1186/s12263-018-0607-5
- Arab L. Biomarkers of fat and fatty acid intake. *J Nutr*. 2003;133(Suppl 3):925S–932S. doi:10.1093/jn/133.3.925S
- Tasevska N, Runswick SA, McTaggart A, et al. Urinary sucrose and fructose as biomarkers for sugar consumption. *Cancer Epidemiol Biomarkers Prev*. 2005;14:1287–1294. doi:10.1158/1055-9965.Epi-04-0827
- Michielsen C, Almanza-Aguilera E, Brouwer-Brolsma EM, et al. Biomarkers of food intake for cocoa and liquorice (products): a systematic review. *Genes Nutr*. 2018;13:22. doi:10.1186/s12263-018-0610-x
- Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Academic Press; 1988.
- Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess*. 1994;6:284–290.
- Van der Steen M, Stevens CV. Undecylenic acid: a valuable and physiologically active renewable building block from castor oil. *ChemSusChem*. 2009;2:692–713. doi:10.1002/cssc.200900075
- Vionnet N, Münger LH, Freiburghaus C, et al. Assessment of lactase activity in humans by measurement of galactitol and galactonate in serum and urine after milk intake. *Am J Clin Nutr*. 2019;109:470–477. doi:10.1093/ajcn/nqy296
- Sofie Biong A, Berstad P, Pedersen JI. Biomarkers for intake of dairy fat and dairy products. *Eur J Lipid Sci Technol*. 2006;108:827–834. doi:10.1002/ejlt.200600044
- Baylín A, Kabagambe EK, Siles X, et al. Adipose tissue biomarkers of fatty acid intake. *Am J Clin Nutr*. 2002;76:750–757. doi:10.1093/ajcn/76.4.750
- Brevik A, Veierød MB, Dreven CA, et al. Evaluation of the odd fatty acids 15:0 and 17:0 in serum and adipose tissue as markers of intake of milk and dairy fat. *Eur J Clin Nutr*. 2005;59:1417–1422. doi:10.1038/sj.ejcn.1602256
- Guertin KA, Moore SC, Sampson JN, et al. Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. *Am J Clin Nutr*. 2014;100:208–217. doi:10.3945/ajcn.113.078758

47. Sun Q, Ma J, Campos H, et al. Plasma and erythrocyte biomarkers of dairy fat intake and risk of ischemic heart disease. *Am J Clin Nutr.* 2007;86:929–937. doi:10.1093/ajcn/86.4.929
48. Yakoob MY, Shi P, Hu FB, et al. Circulating biomarkers of dairy fat and risk of incident stroke in U.S. men and women in 2 large prospective cohorts. *Am J Clin Nutr.* 2014;100:1437–1447. doi:10.3945/ajcn.114.083097
49. Yakoob MY, Shi P, Willett WC, et al. Circulating biomarkers of dairy fat and risk of incident diabetes mellitus among men and women in the United States in two large prospective cohorts. *Circulation.* 2016;133:1645–1654. doi:10.1161/circulationaha.115.018410
50. Micha R, King IB, Lemaitre RN, et al. Food sources of individual plasma phospholipid trans fatty acid isomers: the Cardiovascular Health Study. *Am J Clin Nutr.* 2010;91:883–893. doi:10.3945/ajcn.2009.28877
51. Mozaffarian D, de Oliveira Otto MC, Lemaitre RN, et al. *trans*-Palmitoleic acid, other dairy fat biomarkers, and incident diabetes: the Multi-Ethnic Study of Atherosclerosis (MESA). *Am J Clin Nutr.* 2013;97:854–861. doi:10.3945/ajcn.112.045468
52. Wang Y, Hodge RA, Stevens VL, et al. Identification and reproducibility of plasma metabolomic biomarkers of habitual food intake in a US diet validation study. *Metabolites* 2020;10:382. doi:10.3390/metabo10100382
53. Playdon MC, Ziegler RG, Sampson JN, et al. Nutritional metabolomics and breast cancer risk in a prospective study. *Am J Clin Nutr.* 2017;106:637–649. doi:10.3945/ajcn.116.150912
54. Wang Y, Gapstur SM, Carter BD, et al. Untargeted metabolomics identifies novel potential biomarkers of habitual food intake in a cross-sectional study of postmenopausal women. *J Nutr.* 2018;148:932–943. doi:10.1093/jn/nxy027
55. Playdon MC, Sampson JN, Cross AJ, et al. Comparing metabolite profiles of habitual diet in serum and urine. *Am J Clin Nutr.* 2016;104:776–789. doi:10.3945/ajcn.116.135301
56. Smedman AE, Gustafsson IB, Berglund LG, et al. Pentadecanoic acid in serum as a marker for intake of milk fat: relations between intake of milk fat and metabolic risk factors. *Am J Clin Nutr.* 1999;69:22–29. doi:10.1093/ajcn/69.1.22
57. Kotsopoulos J, Tworoger SS, Campos H, et al. Reproducibility of plasma and urine biomarkers among premenopausal and postmenopausal women from the Nurses' Health Studies. *Cancer Epidemiol Biomarkers Prev.* 2010;19:938–946. doi:10.1158/1055-9965.Epi-09-1318
58. Cheung W, Keski-Rahkonen P, Assi N, et al. A metabolomic study of biomarkers of meat and fish intake. *Am J Clin Nutr.* 2017;105:600–608. doi:10.3945/ajcn.116.146639
59. Gibson R, Lau C-HE, Loo RL, et al. The association of fish consumption and its urinary metabolites with cardiovascular risk factors: the International Study of Macro-/Micronutrients and Blood Pressure (INTERMAP). *Am J Clin Nutr.* 2020;111:280–290. doi:10.1093/ajcn/nqz293
60. Stich HF, Hornby AP, Dunn BP. The effect of dietary factors on nitrosoproline levels in human-urine. *Int J Cancer.* 1984;33:625–628. doi:10.1002/ijc.2910330512
61. Olalekan Adeyeye SA, Ashaolu TJ. Heterocyclic amine formation and mitigation in processed meat and meat products: a mini-review. *J Food Prot.* 2021;84:1868–1877. doi:10.4315/jfp-20-471
62. Wedekind R, Keski-Rahkonen P, Robinot N, et al. Syringol metabolites as new biomarkers for smoked meat intake. *Am J Clin Nutr.* 2019;110:1424–1433.
63. Wedekind R, Keski-Rahkonen P, Robinot N, et al. Pepper alkaloids and processed meat intake: results from a randomized trial and the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. *Mol Nutr Food Res.* 2021;65:e2001141. doi:10.1002/mnfr.202001141
64. Abe H, Okuma E, Sekine H, et al. Human urinary excretion of L-histidine-related compounds after ingestion of several meats and fish muscle. *Int J Biochem.* 1993;25:1245–1249.
65. Block WD, Hubbard RW, Steele BF. Excretion of histidine and histidine derivatives by human subjects ingesting protein from different sources. *J Nutr.* 1965;85:419–425.
66. Gil-Agusti M, Esteve-Romero J, Carda-Broch S. Anserine and carnosine determination in meat samples by pure micellar liquid chromatography. *J Chromatogr A.* 2008;1189:444–450. doi:10.1016/j.chroma.2007.11.075
67. Sjolín J, Hjort G, Friman G, et al. Urinary-excretion of 1-methylhistidine—a qualitative indicator of exogenous 3-methylhistidine and intake of meats from various sources. *Metabolism.* 1987;36:1175–1184. doi:10.1016/0026-0495(87)90245-9
68. Shibutani E, Ishii R, Harada S, et al. Charged metabolite biomarkers of food intake assessed via plasma metabolomics in a population-based observational study in Japan. *PLoS One.* 2021;16:e0246456. doi:10.1371/journal.pone.0246456
69. Fraser GE, Jaceldo-Siegl K, Henning SM, et al. Biomarkers of dietary intake are correlated with corresponding measures from repeated dietary recalls and food-frequency questionnaires in the Adventist Health Study-2. *J Nutr.* 2016;146:586–594. doi:10.3945/jn.115.255508
70. Myint T, Fraser GE, Lindsted KD, et al. Urinary 1-methylhistidine is a marker of meat consumption in black and in white California seventh-day Adventists. *10.1093/aje/152.8.752.* *Am J Epidemiol.* 2000;152:752–755.
71. Maitre L, Lau C-HE, Viccaino E, et al. Assessment of metabolic phenotypic variability in children's urine using 1H NMR spectroscopy. *Sci Rep.* 2017;7:46082. doi:10.1038/srep46082 <https://www.nature.com/articles/srep46082#supplementary-information>
72. Xiao Q, Moore SC, Boca SM, et al. Sources of variability in metabolite measurements from urinary samples. *PLoS One.* 2014;9:e95749. doi:10.1371/journal.pone.0095749
73. Sampson JN, Boca SM, Shu XO, et al. Metabolomics in epidemiology: sources of variability in metabolite measurements and implications. *Cancer Epidemiol Biomarkers Prev.* 2013;22:631–640. doi:10.1158/1055-9965.epi-12-1109
74. Townsend MK, Clish CB, Kraft P, et al. Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. *Clin Chem.* 2013;59:1657–1667.
75. Zheng Y, Yu B, Alexander D, et al. Medium-term variability of the human serum metabolome in the Atherosclerosis Risk in Communities (ARIC) study. *Omic.* 2014;18:364–373.
76. Xu L, Sinclair AJ, Faiza M, et al. Furan fatty acids—beneficial or harmful to health? *Prog Lipid Res.* 2017;68:119–137. doi:10.1016/j.plipres.2017.10.002
77. Spiteller G. Furan fatty acids: occurrence, synthesis, and reactions. Are furan fatty acids responsible for the cardioprotective effects of a fish diet? *Lipids.* 2005;40:755–771. doi:10.1007/s11745-005-1438-5
78. Hanhineva K, Lankinen MA, Pedret A, et al. Nontargeted metabolite profiling discriminates diet-specific biomarkers for consumption of whole grains, fatty fish, and bilberries in a randomized controlled trial. *J Nutr.* 2015;145:7–17. doi:10.3945/jn.114.196840
79. Pallister T, Jennings A, Mohny RP, et al. Characterizing blood metabolomics profiles associated with self-reported food intakes in female twins. *PLoS One.* 2016;11:e0158568. doi:10.1371/journal.pone.0158568
80. Zheng Y, Yu B, Alexander D, et al. Human metabolome associates with dietary intake habits among African Americans in the Atherosclerosis Risk in Communities Study. *Am J Epidemiol.* 2014;179:1424–1433. doi:10.1093/aje/kwu073
81. Lloyd AJ, Fave G, Beckmann M, et al. Use of mass spectrometry fingerprinting to identify urinary metabolites after consumption of specific foods. *Am J Clin Nutr.* 2011;94:981–991. doi:10.3945/ajcn.111.017921
82. Andersen M-BS, Reinbach HC, Rinnan Å, et al. Discovery of exposure markers in urine for brassica-containing meals served with different protein sources by UPLC-qTOF-MS untargeted metabolomics. *Metabolomics.* 2013;9:984–997. doi:10.1007/s11306-013-0522-0
83. Agueusop I, Musholt PB, Klaus B, et al. Short-term variability of the human serum metabolome depending on nutritional and metabolic health status. *Sci Rep.* 2020;10:16310. doi:10.1038/s41598-020-72914-7
84. Li-Gao R, Hughes DA, le Cessie S, et al. Assessment of reproducibility and biological variability of fasting and postprandial plasma metabolite concentrations using 1H NMR spectroscopy. *PLoS One.* 2019;14:e0218549. doi:10.1371/journal.pone.0218549
85. Pratić G, Gao Q, Manach C, et al. Biomarkers of food intake for Allium vegetables. *Genes Nutr.* 2018;13:34–34. doi:10.1186/s12263-018-0624-4
86. Brantsæter AL, Haugen M, Rasmussen SE, et al. Urine flavonoids and plasma carotenoids in the validation of fruit, vegetable and tea intake during pregnancy in the Norwegian Mother and Child Cohort Study (MoBa). *Public Health Nutr.* 2007;10:838–847. doi:10.1017/s1368890007339037
87. Campbell DR, Gross MD, Martini MC, et al. Plasma carotenoids as biomarkers of vegetable and fruit intake. *Cancer Epidemiol Biomarkers Prev.* Sep. 1994;3:493–500.
88. Bogers RP, Van Assema P, Kester AD, et al. Reproducibility, validity, and responsiveness to change of a short questionnaire for measuring fruit and vegetable intake. *Am J Epidemiol.* 2004;159:900–909. doi:10.1093/aje/kwh123
89. Marshall JR, Lanza E, Bloch A, et al. Indexes of food and nutrient intakes as predictors of serum concentrations of nutrients: the problem of inadequate discriminant validity. The Polyp Prevention Trial Study Group. *Am J Clin Nutr.* 1997;65:1269s–1274s. doi:10.1093/ajcn/65.4.1269S
90. Jansen MC, Van Kappel AL, Ocké MC, et al. Plasma carotenoid levels in Dutch men and women, and the relation with vegetable and fruit consumption. *Eur J Clin Nutr.* 2004;58:1386–1395. doi:10.1038/sj.ejcn.1601981
91. van Kappel AL, Steghens JP, Zeleniuch-Jacquotte A, et al. Serum carotenoids as biomarkers of fruit and vegetable consumption in the New York Women's Health Study. *Public Health Nutr.* 2001;4:829–835. doi:10.1079/phn2000115
92. Irwig MS, El-Sohemy A, Baylin A, et al. Frequent intake of tropical fruits that are rich in beta-cryptoxanthin is associated with higher plasma beta-cryptoxanthin concentrations in Costa Rican adolescents. *J Nutr.* 2002;132:3161–3167. doi:10.1093/jn/131.10.3161
93. Toft U, Kristoffersen L, Ladelund S, et al. Relative validity of a food frequency questionnaire used in the Inter99 study. *Eur J Clin Nutr.* 2008;62:1038–1046. doi:10.1038/sj.ejcn.1602815
94. Andersen LF, Veierød MB, Johansson L, et al. Evaluation of three dietary assessment methods and serum biomarkers as measures of fruit and vegetable intake, using the method of triads. *Br J Nutr.* 2005;93:519–527. doi:10.1079/bjn20041381
95. Krogholm KS, Bysted A, Brantsæter AL, et al. Evaluation of flavonoids and enterolactone in overnight urine as intake biomarkers of fruits, vegetables and

- beverages in the Inter99 cohort study using the method of triads. *Br J Nutr.* 2012;108:1904–1912. doi:[10.1017/s0007114512000104](https://doi.org/10.1017/s0007114512000104)
96. Mohammadifard N, Omidvar N, Houshiarrad A, et al. Validity and reproducibility of a food frequency questionnaire for assessment of fruit and vegetable intake in Iranian adults. *J Res Med Sci.* 2011;16:1286–1297.
97. Playdon MC, Moore SC, Derkach A, et al. Identifying biomarkers of dietary patterns by using metabolomics. *Am J Clin Nutr.* 2017;105:450–465. doi:[10.3945/ajcn.116.144501](https://doi.org/10.3945/ajcn.116.144501)
98. Sri Harsha PSC, Wahab RA, Garcia-Aloy M, et al. Biomarkers of legume intake in human intervention and observational studies: a systematic review. *Genes Nutr.* 2018;13:25. doi:[10.1186/s12263-018-0614-6](https://doi.org/10.1186/s12263-018-0614-6)
99. Maskarinec G, Singh S, Meng L, et al. Dietary soy intake and urinary isoflavone excretion among women from a multiethnic population. *Cancer Epidemiol Biomarkers Prev.* 1998;7:613–619.
100. Frankenfeld CL, Patterson RE, Horner NK, et al. Validation of a soy food-frequency questionnaire and evaluation of correlates of plasma isoflavone concentrations in postmenopausal women. *Am J Clin Nutr.* 2003;77:674–680. doi:[10.1093/ajcn/77.3.674](https://doi.org/10.1093/ajcn/77.3.674)
101. Zeleniuch-Jacquotte A, Adlercreutz H, Akhmedkhanov A, et al. Reliability of serum measurements of lignans and isoflavonoid phytoestrogens over a two-year period. *Cancer Epidemiol Biomarkers Prev.* 1998;7:885–889.
102. Chukwumah YC, Walker LT, Verghese M, et al. Comparison of extraction methods for the quantification of selected phytochemicals in peanuts (*Arachis hypogaea*). *J Agric Food Chem.* 2007;55:285–290. doi:[10.1021/jf062148t](https://doi.org/10.1021/jf062148t)
103. Kano M, Takayanagi T, Harada K, et al. Bioavailability of isoflavones after ingestion of soy beverages in healthy adults. *J Nutr.* 2006;136:2291–2296. doi:[10.1093/jn/136.9.2291](https://doi.org/10.1093/jn/136.9.2291)
104. King RA, Bursill DB. Plasma and urinary kinetics of the isoflavones daidzein and genistein after a single soy meal in humans. *Am J Clin Nutr.* 1998;67:867–872. doi:[10.1093/ajcn/67.5.867](https://doi.org/10.1093/ajcn/67.5.867)
105. Chávez-Suárez KM, Ortega-Vélez MI, Valenzuela-Quintanar AI, et al. Phytoestrogen concentrations in human urine as biomarkers for dietary phytoestrogen intake in Mexican women. *Nutrients* 2017;9:1078. doi:[10.3390/nu9101078](https://doi.org/10.3390/nu9101078)
106. Kim J, Kim HJ, Joong H, et al. Overnight urinary excretion of isoflavones as an indicator for dietary isoflavone intake in Korean girls of pubertal age. *Br J Nutr.* 2010;104:709–715. doi:[10.1017/s0007114510000978](https://doi.org/10.1017/s0007114510000978)
107. Tseng M, Olufade T, Kurzer MS, et al. Food frequency questionnaires and overnight urines are valid indicators of daidzein and genistein intake in U.S. women relative to multiple 24-h urine samples. *Nutr Cancer.* 2008;60:619–626. doi:[10.1080/01635580801993751](https://doi.org/10.1080/01635580801993751)
108. Whittom C, Neelakantan N, Ong CN, et al. Reproducibility of dietary biomarkers in a multiethnic Asian population. *Mol Nutr Food Res.* 2019;63:e1801104. doi:[10.1002/mnfr.201801104](https://doi.org/10.1002/mnfr.201801104)
109. Frankenfeld CL. O-desmethylandrogens: the importance of equol's lesser known cousin to human health. *Adv Nutr.* 2011;2:317–324. doi:[10.3945/an.111.000539](https://doi.org/10.3945/an.111.000539)
110. Lipovac M, Pfitscher A, Hobiger S, et al. Red clover isoflavone metabolite bioavailability is decreased after fructooligosaccharide supplementation. *Fitoterapia.* 2015;105:93–101. doi:[10.1016/j.fitote.2015.06.011](https://doi.org/10.1016/j.fitote.2015.06.011)
111. Fujioka N, Ransom BW, Carmella SG, et al. Harnessing the power of cruciferous vegetables: developing a biomarker for brassica vegetable consumption using urinary 3,3'-diindolylmethane. *Cancer Prev Res (Phila).* 2016;9:788–793. doi:[10.1158/1940-6207.Ccrp-16-0136](https://doi.org/10.1158/1940-6207.Ccrp-16-0136)
112. S C Sri Harsha P, Abdul Wahab R, Cuparencu C, et al. A metabolomics approach to the identification of urinary biomarkers of pea intake. *Nutrients.* 2018;10:1911. doi:[10.3390/nu10121911](https://doi.org/10.3390/nu10121911)
113. Malik VS, Guasch-Ferre M, Hu FB, et al. Identification of plasma lipid metabolites associated with nut consumption in US men and women. *J Nutr.* 2019;149:1215–1221. doi:[10.1093/jn/nxz048](https://doi.org/10.1093/jn/nxz048)
114. Perera T, Young MR, Zhang Z, et al. Identification and monitoring of metabolite markers of dry bean consumption in parallel human and mouse studies. *Mol Nutr Food Res.* 2015;59:795–806. doi:[10.1002/mnfr.201400847](https://doi.org/10.1002/mnfr.201400847)
115. Lang R, Lang T, Bader M, et al. High-throughput quantitation of proline betaine in foods and suitability as a valid biomarker for citrus consumption. *J Agric Food Chem.* 2017;65:1613–1619. doi:[10.1021/acs.jafc.6b05824](https://doi.org/10.1021/acs.jafc.6b05824)
116. Escarpa A, González MC. High-performance liquid chromatography with diode-array detection for the determination of phenolic compounds in peel and pulp from different apple varieties. *J Chromatogr A.* 1998;823:331–337. doi:[10.1016/s0021-9673\(98\)00294-5](https://doi.org/10.1016/s0021-9673(98)00294-5)
117. Tsao R, Yang R, Young JC, et al. Polyphenolic profiles in eight apple cultivars using high-performance liquid chromatography (HPLC). *J Agric Food Chem.* 2003;51:6347–6353. doi:[10.1021/jf0346298](https://doi.org/10.1021/jf0346298)
118. Kanazawa K, Sakakibara H. High content of dopamine, a strong antioxidant, in Cavendish banana. *J Agric Food Chem.* 2000;48:844–848. doi:[10.1021/jf9909860](https://doi.org/10.1021/jf9909860)
119. Harnly JM, Doherty RF, Beecher GR, et al. Flavonoid content of U.S. fruits, vegetables, and nuts. *J Agric Food Chem.* 2006;54:9966–9977. doi:[10.1021/jf061478a](https://doi.org/10.1021/jf061478a)
120. Wingerath T, Stahl W, Sies H. beta-Cryptoxanthin selectively increases in human chylomicrons upon ingestion of tangerine concentrate rich in beta-cryptoxanthin esters. *Arch Biochem Biophys.* 1995;324:385–390. doi:[10.1006/abbi.1995.0052](https://doi.org/10.1006/abbi.1995.0052)
121. Khoo HE, Prasad KN, Kong KW, et al. Carotenoids and their isomers: color pigments in fruits and vegetables. *Molecules.* 2011;16:1710–1738. doi:[10.3390/molecules16021710](https://doi.org/10.3390/molecules16021710)
122. Ferrari P, Al-Delaimy WK, Slimani N, et al. An approach to estimate between- and within-group correlation coefficients in multicenter studies: plasma carotenoids as biomarkers of intake of fruits and vegetables. *Am J Epidemiol.* 2005;162:591–598. doi:[10.1093/aje/kwi242](https://doi.org/10.1093/aje/kwi242)
123. Cena H, Roggi C, Turconi G. Development and validation of a brief food frequency questionnaire for dietary lutein and zeaxanthin intake assessment in Italian women. *Eur J Nutr.* 2008;47:1–9. doi:[10.1007/s00394-007-0689-0](https://doi.org/10.1007/s00394-007-0689-0)
124. Zheng JS, Sharp SJ, Imamura F, et al. Association of plasma biomarkers of fruit and vegetable intake with incident type 2 diabetes: EPIC-InterAct case-cohort study in eight European countries. *BMJ.* 2020;370:m2194. doi:[10.1136/bmj.m2194](https://doi.org/10.1136/bmj.m2194)
125. Erlund I, Meririnne E, Alfthan G, et al. Plasma kinetics and urinary excretion of the flavanones naringenin and hesperetin in humans after ingestion of orange juice and grapefruit juice. *J Nutr.* 2001;131:235–241. doi:[10.1093/jn/131.2.235](https://doi.org/10.1093/jn/131.2.235)
126. Vazquez-Manjarrez N, Weinert CH, Ulaszewska MM, et al. Discovery and validation of banana intake biomarkers using untargeted metabolomics in human intervention and cross-sectional studies. *J Nutr.* 2019;149:1701–1713. doi:[10.1093/jn/nxz125](https://doi.org/10.1093/jn/nxz125)
127. Pujos-Guillot E, Hubert J, Martin JF, et al. Mass spectrometry-based metabolomics for the discovery of biomarkers of fruit and vegetable intake: citrus fruit as a case study. *J Proteome Res.* 2013;12:1645–1659. doi:[10.1021/pr300997c](https://doi.org/10.1021/pr300997c)
128. Gibbons H, Michielsen CJR, Rundle M, et al. Demonstration of the utility of biomarkers for dietary intake assessment; proline betaine as an example. *Mol Nutr Food Res.* 2017;61:1700037. doi:[10.1002/mnfr.201700037](https://doi.org/10.1002/mnfr.201700037)
129. Brevik A, Rasmussen SE, Drevon CA, et al. Urinary excretion of flavonoids reflects even small changes in the dietary intake of fruits and vegetables. *Cancer Epidemiol Biomarkers Prev.* 2004;13:843–849.
130. McNamara AE, Collins C, Harsha P, et al. Metabolomic-based approach to identify biomarkers of apple intake. *Mol Nutr Food Res.* 2020;64:e1901158. doi:[10.1002/mnfr.201901158](https://doi.org/10.1002/mnfr.201901158)
131. Garcia-Aloy M, Llorach R, Urpi-Sarda M, et al. Nutrimetabolomics fingerprinting to identify biomarkers of bread exposure in a free-living population from the PREDIMED study cohort. *Metabolomics.* 2015;11:155–165. doi:[10.1007/s11306-014-0682-6](https://doi.org/10.1007/s11306-014-0682-6)
132. Kärkkäinen O, Lankinen MA, Vitale M, et al. Diets rich in whole grains increase betainized compounds associated with glucose metabolism. *Am J Clin Nutr.* 2018;108:971–979. doi:[10.1093/ajcn/nqy169](https://doi.org/10.1093/ajcn/nqy169)
133. Nordin E, Steffensen SK, Laursen BB, et al. An inverse association between plasma benzoxazinoid metabolites and PSA after rye intake in men with prostate cancer revealed with a new method. *Sci Rep.* 2022;12:5260. doi:[10.1038/s41598-022-08856-z](https://doi.org/10.1038/s41598-022-08856-z)
134. Kozubek A, Tyman JH. Resorcinolic lipids, the natural non-isoprenoid phenolic amphiphiles and their biological activity. *Chem Rev.* 1999;99:1–26. doi:[10.1021/cr970464o](https://doi.org/10.1021/cr970464o)
135. Chen Y, Ross AB, Aman P, et al. Alkylresorcinols as markers of whole grain wheat and rye in cereal products. *J Agric Food Chem.* 2004;52:8242–8246. doi:[10.1021/jf049726v](https://doi.org/10.1021/jf049726v)
136. Koistinen VM, Mattila O, Katina K, et al. Metabolic profiling of sourdough fermented wheat and rye bread. *Sci Rep.* 2018;8:5684. doi:[10.1038/s41598-018-24149-w](https://doi.org/10.1038/s41598-018-24149-w)
137. Ross AB, Shepherd MJ, Schüpphaus M, et al. Alkylresorcinols in cereals and cereal products. *J Agric Food Chem.* 2003;51:4111–4118. doi:[10.1021/jf0340456](https://doi.org/10.1021/jf0340456)
138. Landberg R, Aman P, Friberg LE, et al. Dose response of whole-grain biomarkers: alkylresorcinols in human plasma and their metabolites in urine in relation to intake. *Am J Clin Nutr.* 2009;89:290–296. doi:[10.3945/ajcn.2008.26709](https://doi.org/10.3945/ajcn.2008.26709)
139. Söderholm PP, Lundin JE, Koskela AH, et al. Pharmacokinetics of alkylresorcinol metabolites in human urine. *Br J Nutr.* 2011;106:1040–1044. doi:[10.1017/s0007114511001383](https://doi.org/10.1017/s0007114511001383)
140. Zhu Y, Shuriknight KL, Chen X, et al. Identification and pharmacokinetics of novel alkylresorcinol metabolites in human urine, new candidate biomarkers for whole-grain wheat and rye intake. *J Nutr.* 2014;144:114–122. doi:[10.3945/jn.113.184663](https://doi.org/10.3945/jn.113.184663)
141. Montonen J, Landberg R, Kamal-Eldin A, et al. Reliability of fasting plasma alkylresorcinol concentrations measured 4 months apart. *Eur J Clin Nutr.* 2010;64:698–703. doi:[10.1038/ejcn.2010.71](https://doi.org/10.1038/ejcn.2010.71)
142. Montonen J, Landberg R, Kamal-Eldin A, et al. Reliability of fasting plasma alkylresorcinol metabolites concentrations measured 4 months apart. *Eur J Clin Nutr.* 2012;66:968–970. doi:[10.1038/ejcn.2012.66](https://doi.org/10.1038/ejcn.2012.66)
143. Landberg R, Aman P, Hallmans G, et al. Long-term reproducibility of plasma alkylresorcinols as biomarkers of whole-grain wheat and rye intake within Northern Sweden Health and Disease Study Cohort. *Eur J Clin Nutr.* 2013;67:259–263. doi:[10.1038/ejcn.2013.10](https://doi.org/10.1038/ejcn.2013.10)
144. Landberg R, Wierzbicka R, Shi L, et al. New alkylresorcinol metabolites in spot urine as biomarkers of whole grain wheat and rye intake in a Swedish middle-aged population. *Eur J Clin Nutr.* 2018;72:1439–1446. doi:[10.1038/s41430-017-0079-5](https://doi.org/10.1038/s41430-017-0079-5)

145. Wierzbicka R, Zamaratskaia G, Kamal-Eldin A, et al. Novel urinary alkylresorcinol metabolites as biomarkers of whole grain intake in free-living Swedish adults. *Mol Nutr Food Res*. 2017;61:1700015. doi:10.1002/mnfr.201700015
146. Sang S, Chu Y. Whole grain oats, more than just a fiber: role of unique phytochemicals. *Mol Nutr Food Res*. 2017;61:1600715. doi:10.1002/mnfr.201600715
147. Wang P, Zhang S, Yerke A, et al. Avenanthramide metabolite from whole-grain oat intake is influenced by *Faecalibacterium prausnitzii* in healthy adults. *J Nutr*. 2021;151:1426–1435. doi:10.1093/jn/nxab006
148. Zhang T, Shao J, Gao Y, et al. Absorption and elimination of oat avenanthramides in humans after acute consumption of oat cookies. *Oxid Med Cell Longev*. 2017;2017:2056705. doi:10.1155/2017/2056705
149. Hanhineva K, Brunius C, Andersson A, et al. Discovery of urinary biomarkers of whole grain rye intake in free-living subjects using nontargeted LC-MS metabolite profiling. *Mol Nutr Food Res*. 2015;59:2315–2325. doi:10.1002/mnfr.201500423
150. Kuhnle GG, Joosen AM, Wood TR, et al. Detection and quantification of sucrose as dietary biomarker using gas chromatography and liquid chromatography with mass spectrometry. *Rapid Commun Mass Spectrom*. 2008;22:279–282. doi:10.1002/rcm.3355
151. Moore LB, Liu SV, Halliday TM, et al. Urinary excretion of sodium, nitrogen, and sugar amounts are valid biomarkers of dietary sodium, protein, and high sugar intake in nonobese adolescents. *J Nutr*. 2017;147:2364–2373. doi:10.3945/jn.117.256875
152. Valenzuela LO, O'Grady SP, Enright LE, et al. Evaluation of childhood nutrition by dietary survey and stable isotope analyses of hair and breath. *Am J Hum Biol*. 2018;30:e23103. doi:10.1002/ajhb.23103
153. Yun HY, Tinker LF, Neuhauser ML, 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. doi:10.1093/jn/nxaa195
154. Muli S, Goerdten J, Oluwagbemigun K, et al. A systematic review of metabolomic biomarkers for the intake of sugar-sweetened and low-calorie sweetened beverages. *Metabolites*. 2021;11:546. doi:10.3390/metabo11080546
155. Ramne S, Gray N, Hellstrand S, et al. Comparing self-reported sugar intake with the sucrose and fructose biomarker from overnight urine samples in relation to cardiometabolic risk factors. *Front Nutr*. 2020;7:62. doi:10.3389/fnut.2020.00062
156. Abreu TC, Hulshof PJM, Boshuizen HC, et al. Validity coefficient of repeated measurements of urinary marker of sugar intake is comparable to urinary nitrogen as marker of protein intake in free-living subjects. *Cancer Epidemiol Biomarkers Prev*. 2021;30:193–202. doi:10.1158/1055-9965.Epi-20-0271
157. MacDougall CR, Hill CE, Jähren AH, et al. The $\delta^{13}\text{C}$ 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. doi:10.1093/jn/nxx017
158. Beckmann M, Joosen AM, Clarke MM, et al. Changes in the human plasma and urinary metabolome associated with acute dietary exposure to sucrose and the identification of potential biomarkers of sucrose intake. *Mol Nutr Food Res*. 2016;60:444–457. doi:10.1002/mnfr.201500495
159. Le MT, Frye RF, Rivard CJ, et al. Effects of high-fructose corn syrup and sucrose on the pharmacokinetics of fructose and acute metabolic and hemodynamic responses in healthy subjects. *Metabolism*. 2012;61:641–651.
160. Votruba SB, Shaw PA, Oh EJ, et al. 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. doi:10.1093/ajcn/nqz208
161. Quifer-Rada P, Chiva-Blanch G, Jauregui O, et al. A discovery-driven approach to elucidate urinary metabolome changes after a regular and moderate consumption of beer and nonalcoholic beer in subjects at high cardiovascular risk. *Mol Nutr Food Res*. 2017;61:1600980. doi:10.1002/mnfr.201600980
162. Quifer-Rada P, Martinez-Huelamo M, Chiva-Blanch G, et al. Urinary isoxanthohumol is a specific and accurate biomarker of beer consumption. *J Nutr*. 2014;144:484–488. doi:10.3945/jn.113.185199
163. Gurdeniz G, Jensen MG, Meier S, et al. Detecting beer intake by unique metabolite patterns. *J Proteome Res*. 2016;15:4544–4556. doi:10.1021/acs.jproteome.6b00635
164. van Breemen RB, Yuan Y, Banuvar S, et al. Pharmacokinetics of prenylated hop phenols in women following oral administration of a standardized extract of hops. *Mol Nutr Food Res*. 2014;58:1962–1969. doi:10.1002/mnfr.201400245
165. Esteban-Fernandez A, Ibanez C, Simo C, et al. An ultrahigh-performance liquid chromatography-time-of-flight mass spectrometry metabolomic approach to studying the impact of moderate red-wine consumption on urinary metabolome. *J Proteome Res*. 2018;17:1624–1635. doi:10.1021/acs.jproteome.7b00904
166. Liu SQ. Malolactic fermentation in wine—beyond deacidification. *J Appl Microbiol*. 2002;92:589–601. doi:10.1046/j.1365-2672.2002.01589.x
167. Son H-S, Kim KM, Van den Berg F, et al. H-1 nuclear magnetic resonance-based metabolomic characterization of wines by grape varieties and production areas. *J Agric Food Chem*. 2008;56:8007–8016. doi:10.1021/jf801424u
168. Son H-S, Hwang G-S, Kim KM, et al. Metabolomic studies on geographical grapes and their wines using H-1 NMR analysis coupled with multivariate statistics. *J Agric Food Chem*. 2009;57:1481–1490. Feb25 doi:10.1021/jf803388w
169. Neveu V, Perez-Jimenez J, Vos F, et al. Phenol-explorer: an online comprehensive database on polyphenol contents in foods. *Database (Oxford)*. 2010;2010:bap024. doi:10.1093/database/bap024
170. Zamora-Ros R, Urpi-Sarda M, Lamuela-Raventos RM, et al. Resveratrol metabolites in urine as a biomarker of wine intake in free-living subjects: the PREDIMED study. *Free Radic Biol Med*. 2009;46:1562–1566. doi:10.1016/j.freeradbiomed.2008.12.023
171. Zamora-Ros R, Urpi-Sarda M, Lamuela-Raventos RM, et al. Diagnostic performance of urinary resveratrol metabolites as a biomarker of moderate wine consumption. *Clin Chem*. 2006;52:1373–1380. doi:10.1373/clinchem.2005.065870
172. Regueiro J, Vallverdu-Queralt A, Simal-Gandara J, et al. Development of a LC-ESI-MS/MS approach for the rapid quantification of main wine organic acids in human urine. *J Agric Food Chem*. 2013;61:6763–6768. doi:10.1021/jf401839g
173. Shahrazad S, Aoyagi K, Winter A, et al. Pharmacokinetics of gallic acid and its relative bioavailability from tea in healthy humans. *J Nutr*. 2001;131:1207–1210.
174. van der Pijl PC, Foltz M, Glube ND, et al. Pharmacokinetics of black tea-derived phenolic acids in plasma. *J Funct Foods*. 2015;17:667–675. doi:10.1016/j.jff.2015.06.020
175. Walle T, Hsieh F, DeLegge MH, et al. High absorption but very low bioavailability of oral resveratrol in humans. *Drug Metab Dispos*. 2004;32:1377–1382. doi:10.1124/dmd.104.000885
176. Landberg R. Can urinary ethyl glucuronide be used as a biomarker of habitual alcohol consumption? *J Nutr*. 2019;149:2077–2078. doi:10.1093/jn/nxz201
177. Walsham NE, Sherwood RA. Ethyl glucuronide and ethyl sulfate. *Adv Clin Chem*. 2014;67:47–71. doi:10.1016/bs.acc.2014.09.006
178. Gnann H, Weinmann W, Thierauf A. Formation of phosphatidylethanol and its subsequent elimination during an extensive drinking experiment over 5 days. *Alcohol Clin Exp Res*. 2012;36:1507–1511. doi:10.1111/j.1530-0277.2012.01768.x
179. Allen NE, Grace PB, Ginn A, 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*. 2008;99:653–659. doi:10.1017/s000711450782407x
180. de Oliveira Otto MC, Nettleton JA, Lemaitre RN, et al. Biomarkers of dairy fatty acids and risk of cardiovascular disease in the Multi-ethnic Study of Atherosclerosis. *J Am Heart Assoc*. 2013;2:e000092. doi:10.1161/jaha.113.000092
181. Lucas M, Proust F, Blanchet C, et al. Is marine mammal fat or fish intake most strongly associated with omega-3 blood levels among the Nunavik Inuit? *Prostaglandins Leukot Essent Fatty Acids*. 2010;83:143–150. doi:10.1016/j.plefa.2010.06.006
182. Zhao C-N, Tang G-Y, Cao S-Y, et al. Phenolic profiles and antioxidant activities of 30 tea infusions from green, black, oolong, white, yellow and dark teas. *Antioxidants (Basel)*. 2019;8:215. doi:10.3390/antiox8070215
183. Mennen LI, Sapinho D, Ito H, et al. Urinary excretion of 13 dietary flavonoids and phenolic acids in free-living healthy subjects—variability and possible use as biomarkers of polyphenol intake. *Eur J Clin Nutr*. 2008;62:519–525. doi:10.1038/sj.ejcn.1602744
184. Edmonds WM, Ferrari P, Rothwell JA, et al. Polyphenol metabolome in human urine and its association with intake of polyphenol-rich foods across European countries. *Am J Clin Nutr*. 2015;102:905–913. doi:10.3945/ajcn.114.101881
185. Hodgson JM, Chan SY, Puddey IB, et al. Phenolic acid metabolites as biomarkers for tea- and coffee-derived polyphenol exposure in human subjects. *Br J Nutr*. 2004;91:301–306. doi:10.1079/bjn20031046
186. Ludwig IA, Clifford MN, Lean MEJ, et al. Coffee: biochemistry and potential impact on health. *Food Funct*. 2014;5:1695–1717. doi:10.1039/c4fo00042k
187. Guertin KA, Loftfield E, Boca SM, et al. Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cancer. *Am J Clin Nutr*. 2015;101:1000–1011. doi:10.3945/ajcn.114.096099
188. Rothwell JA, Keski-Rahkonen P, Robinot N, et al. A metabolomic study of biomarkers of habitual coffee intake in four European countries. *Mol Nutr Food Res*. 2019;63:e1900659.
189. Middtun Ø, Ulvik A, Nygård O, et al. Performance of plasma trigonelline as a marker of coffee consumption in an epidemiologic setting. *Am J Clin Nutr*. 2018;107:941–947. doi:10.1093/ajcn/nqy059
190. Lenk G, Ullah S, Stemme G, et al. Evaluation of a volumetric dried blood spot card using a gravimetric method and a bioanalytical method with capillary blood from 44 volunteers. *Anal Chem*. 2019;91:5558–5565. doi:10.1021/acs.analchem.8b02905
191. Couillard C, Lemieux S, Vohl M-C, et al. Carotenoids as biomarkers of fruit and vegetable intake in men and women. *Br J Nutr*. 2016;116:1206–1215. doi:10.1017/S0007114516003056
192. Picó C, Serra F, Rodríguez AM, et al. Biomarkers of nutrition and health: new tools for new approaches. *Nutrients*. 2019;11:1092. doi:10.3390/nu11051092

193. Vanaelst B, Huybrechts I, Michels N, et al. Mineral concentrations in hair of Belgian elementary school girls: reference values and relationship with food consumption frequencies. *Biol Trace Elem Res*. 2012;150:56–67. doi:[10.1007/s12011-012-9495-5](https://doi.org/10.1007/s12011-012-9495-5)
194. Scalbert A, Rothwell JA, Keski-Rahkonen PVN. The food metabolome and dietary biomarkers. In: Schoeller DA, Westerterp M, eds. *Advances in the Assessment of Dietary Intake*. CRC Press; 2017:259–282.
195. Witwer KW, Zhang CY. Diet-derived microRNAs: unicorn or silver bullet? *Genes Nutr*. 2017;12:15. doi:[10.1186/s12263-017-0564-4](https://doi.org/10.1186/s12263-017-0564-4)
196. Beckmann M, Wilson T, Zubair H, et al. A standardized strategy for simultaneous quantification of urine metabolites to validate development of a biomarker panel allowing comprehensive assessment of dietary exposure. *Mol Nutr Food Res*. 2020;64:e2000517. doi:[10.1002/mnfr.202000517](https://doi.org/10.1002/mnfr.202000517)
197. Noerman S, Landberg R. Healthy dietary patterns reflected by the blood metabolome. *J Intern Med*. 2023;293:408–432.
198. D'Angelo S, Gormley IC, McNulty BA, et al. Combining biomarker and food intake data: calibration equations for citrus intake. *Am J Clin Nutr*. 2019;110:977–983. doi:[10.1093/ajcn/nqz168](https://doi.org/10.1093/ajcn/nqz168)
199. Knudsen MD, Kyrø C, Olsen A, et al. Self-reported whole-grain intake and plasma alkylresorcinol concentrations in combination in relation to the incidence of colorectal cancer. *Am J Epidemiol*. 2014;179:1188–1196. doi:[10.1093/aje/kwu031](https://doi.org/10.1093/aje/kwu031)