The food metabolome: a window over dietary exposure[1–3]

Augustin Scalbert, Lorraine Brennan, Claudine Manach, Cristina Andres-Lacueva, Lars O Dragsted, John Draper, Stephen M Rappaport, Justin JJ van der Hooft, and David S Wishart

ABSTRACT

The food metabolome is defined as the part of the human metabolome directly derived from the digestion and biotransformation of foods and their constituents. With >25,000 compounds known in various foods, the food metabolome is extremely complex, with a composition varying widely according to the diet. By its very nature it represents a considerable and still largely unexploited source of novel dietary biomarkers that could be used to measure dietary exposures with a high level of detail and precision. Most dietary biomarkers currently have been identified on the basis of our knowledge of food compositions by using hypothesis-driven approaches. However, the rapid development of metabolomics resulting from the development of highly sensitive modern analytic instruments, the availability of metabolite databases, and progress in (bio)informatics has made agnostic approaches more attractive as shown by the recent identification of novel biomarkers of intakes for fruit, vegetables, beverages, meats, or complex diets. Moreover, examples also show how the scrutiny of the food metabolome can lead to the discovery of bioactive molecules and dietary factors associated with diseases. However, researchers still face hurdles, which slow progress and need to be resolved to bring this emerging field of research to maturity. These limits were discussed during the First International Workshop on the Food Metabolome held in Glasgow. Key recommendations made during the workshop included more coordination of efforts; development of new databases, software tools, and chemical libraries for the food metabolome; and shared repositories of metabolomic data. Once achieved, major progress can be expected toward a better understanding of the complex interactions between diet and human health. *Am J Clin Nutr* 2014;99:1286–308.

INTRODUCTION

The 2 major achievements of nutrition research in the 20th century were the discovery of essential nutrients and the elucidation of their role in key physiologic functions. Recommendations were defined to provide adequate intakes of these nutrients that led to reduction in risks of deficiency diseases, at least in high-income Western societies. The past 2 decades have seen a shift in nutrition research away from the prevention of deficiency diseases toward the prevention of chronic diseases and the elucidation of the role of nonessential food constituents on such diseases (1). This constitutes a considerable challenge for nutrition research in the 21st century, in particular because of the extreme variety of these bioactive constituents and the large diversity of biochemical targets and signaling and metabolic pathways they may interact with.

Although classical hypothesis-driven approaches have been very successful in discovering essential nutrients, they are ill adapted to aid our understanding of the role of highly diverse nonessential compounds in foods. Data-driven approaches and "omics" technologies offer opportunities to explore the complex interactions between diet and the human organism. In particular, the measurement of hundreds or thousands of metabolites in metabolomic experiments now allows the characterization of individual phenotypes with a level of precision never before achieved (2). Individuals or populations exposed to different environments, lifestyles, or diets can be distinguished and characteristic metabolic differences identified (3).

A growing number of metabolomic studies have been published over the past 5 y in the field of nutrition (3–6). Metabolomics was used to show the alteration of metabolic profiles on the consumption of specific nutrients, foods, or diets in small-scale intervention studies. Two different fractions of the human metabolome are influenced by the diet: the endogenous metabolome and the food metabolome (**Figure 1**). The endogenous metabolome includes all metabolites from the host. Its variations show novel metabolic effects of the diet that may affect human health. The "food metabolome" has been defined as the sum of

[1] From the International Agency for Research on Cancer, Lyon, France (AS); University College Dublin, Dublin, Ireland (LB); the Institut National de la Recherche Agronomique, Clermont-Ferrand, France (CM); Clermont University, Clermont-Ferrand, France (CM); the University of Barcelona, Barcelona, Spain (CA-L); the University of Copenhagen, Frederiksberg, Denmark (LOD); Aberystwyth University, Aberystwyth, United Kingdom (JD); the University of California, Berkeley, CA (SMR); the University of Glasgow, Glasgow, United Kingdom (JJJvdH); and the University of Alberta, Edmonton, Canada (DSW).

[2] Supported by the European Union (NutriTech FP7-KBBE-2011-5 grant 289511, EUROCAN FP7-KBBE-2010.2.4.1-2 grant 260791); the Danish Ministry of Science, Technology, and Innovation (for the UNIK Food, Fitness and Pharma Project); the French National Agency for Research (PhenoMeNEp ANR-10-ALIA-007); the Medical Research Council (MR/J010308/1); and the Spanish Ministerio de Economia y Competitividad (MINECO; project AGL2009-13906-C02-01) and by a Senior Visiting Scientist Award (to SMR) granted by the International Agency for Research on Cancer.

[3] Address correspondence to A Scalbert, International Agency for Research on Cancer (IARC), Nutrition and Metabolism Section, Biomarkers Group, 150 Cours Albert Thomas, F-69372 Lyon Cedex 08, France. E-mail: scalberta@iarc.fr.

Received September 19, 2013. Accepted for publication March 24, 2014.
First published online April 23, 2014; doi: 10.3945/ajcn.113.076133.

Supplemental Material can be found at:
http://ajcn.nutrition.org/content/suppl/2014/05/13/ajcn.113.076133.DCSupplemental.html

all metabolites directly derived from the digestion of foods, their absorption in the gut, and biotransformation by the host tissues and the microbiota as first proposed by one of the authors of this review in 2008 (7). Other authors have also proposed to define the "food metabolome" as the whole set of food constituents in any foods (5, 8). A definition of the metabolome centered on biological species is preferred here. Humans consume as many metabolomes as there are biological species making up our foods—for example, the tomato or beef metabolomes. Therefore, the human metabolome contains fractions of these metabolomes, partly transformed after ingestion, which constitute the human food metabolome.

The various foods consumed by humans contain >25,000 compounds, most of them being further metabolized in the human body (9). The food metabolome is therefore highly complex and also highly variable. This variability constitutes a unique and extremely rich source of information on the human diet that has barely been exploited. Detailed characterization of the food metabolome should permit accurate monitoring of dietary exposure and identification of foods that influence disease risks in clinical and epidemiologic studies. This review describes the current knowledge on the food metabolome and discusses opportunities for nutrition research. It also makes recommendations to move the field forward as discussed by the participants in the First International Workshop on the Food Metabolome (4–5 June 2013, Glasgow, United Kingdom), which convened with 50 experts from Europe and North America (Supplemental Table 1 under "Supplemental data" in the online issue).

DIETARY BIOMARKERS IN THE PRE-OMICS ERA

Studies of connections between the diet or specific dietary factors and health status require accurate measurements of dietary exposures. Such measurements can be used to evaluate compliance in dietary intervention studies, to find associations with disease outcomes, or to monitor dietary changes in populations. Dietary exposure has traditionally been measured with self-reported methods, namely dietary recalls or food-frequency questionnaires (10). However, a number of random and systematic errors are inherent in such methods, including recall bias and difficulty in assessing portion sizes (11). The resulting misclassification of subjects, especially when sorting them according to dietary intake, can influence observed associations between dietary exposures and disease outcomes and underlies inconsistencies in published findings in the field of nutritional epidemiology (12).

To address these shortcomings, intense efforts have been directed toward statistical techniques to correct measurement errors as well as toward developing new dietary assessment instruments. The application of dietary biomarkers as more objective measures of dietary exposure in nutritional epidemiology has been particularly significant (13). These biomarkers have been used as measures of nutritional status and of exposure to bioactive molecules in foods, as surrogate indicators of food intake, and to validate measures of dietary intake (14). Biomarkers are also useful when little or no data exist on food composition, as is often the case for bioactive molecules such as glucosinolates or food contaminants such as aflatoxins (15, 16).

Dietary biomarkers measured in population studies

A variety of dietary biomarkers identified through the analysis of correlations with dietary intake have been measured in epidemiologic studies. Information on these biomarkers has been systematically collected in the novel Exposome-Explorer database (V Neveu, DS Wishart, and A Scalbert, unpublished data, 2014); ~100 biomarkers could be identified (Supplemental Table 2 under "Supplemental data" in the online issue). These biomarkers have been measured in plasma or serum (carotenoids, fatty acids, vitamins, polyphenols, food contaminants, and

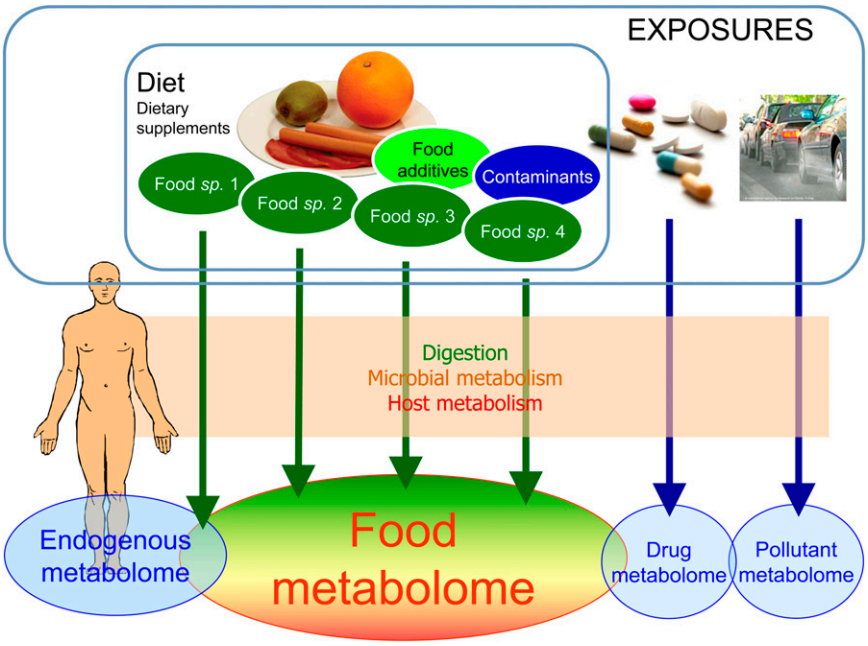

FIGURE 1. The human metabolomes. *sp.*, species.

enzymes), red blood cells (fatty acids, carotenoids, and hemoglobin adducts), and to a lesser extent in urine (polyphenols, vitamins, inorganic compounds, and amino acids). Some of these biomarkers correspond to nutrients and bioactive compounds and have been used to compare status or exposure. Some have been used as surrogate biomarkers of food intake, as follows: polyphenols, carotenoids, and vitamin C for fruit and vegetables (17, 18); alkylresorcinols for whole-grain cereals (19, 20); isoflavones for soy (21); amino acids and fatty acids for meat (22, 23); fatty acids for dairy products and fish (22, 24); and polyphenols for tea and wine (18, 25) (**Table 1**). Dietary biomarkers not only include natural food constituents but also certain food additives such as iodine in milk (26) or food contaminants such as polychlorinated biphenyls in fatty fish (27). These latter biomarkers are often specific to certain populations who consume these additives or where consistent levels of contamination are observed.

Other biomarkers are directly derived from the digestion and gut absorption of food constituents or are endogenous metabolites that have been altered by exposure to specific nutrients. For instance, serotonin metabolism is altered by acute alcohol intake (28), the activity of selenium-containing enzymes such as erythrocyte glutathione peroxidase depends on selenium intake, and ceramide synthase is inhibited by exposure to the mycotoxin fumonisins (29).

Pharmacokinetics and reliability of dietary biomarkers

Dietary biomarkers are not without their limitations. They may be altered because of possible interactions with genetic factors, physiologic or health status (ie, age or obesity) (30), dietary factors such as fats for lipophilic biomarkers (31), and lifestyle factors such as alcohol intake or smoking (32). Their concentrations also vary over time according to their pharmacokinetic properties. A higher intraindividual variability is expected for biomarkers with a short half-life (20, 33). Intraindividual variability leads to exposure measurement errors when the objective is to characterize habitual exposure in epidemiologic studies and small numbers of measurements are available across subjects.

Some of the biomarkers listed in Table 1 have half-lives that do not exceed 24 h [polyphenols, alkylresorcinols, and amino acids (34, 35)]. These biomarkers may thus be useful only in populations who regularly and frequently consume these dietary sources. Lipophilic markers (carotenoids, lipids) (36) or biomarkers associated with erythrocytes (folate, fatty acids) (29) have longer half-lives (week to month) because of the equilibrium of biomarkers between blood and fatty tissues, or because of their integration into erythrocytes. Some dietary compounds such as isothiocyanates and acrylamide also form adducts with blood albumin and hemoglobin (37, 38), with half-lives varying between 3 and 8 wk, and may be used as longer-term biomarkers. Protein adducts with dietary compounds have received limited attention thus far. Adductomics appears to be particularly promising for the discovery of these adduct biomarkers (39, 40).

Biomarker sensitivity and specificity

Dietary biomarkers should have sufficient sensitivity to measure exposures within ranges commonly found in the populations of interest. Intervention studies are essential to address this question and to evaluate the relation between exposure and biomarker concentrations (17, 41). Biomarkers such as vitamin C or selenium in erythrocyte glutathione peroxidase show saturable effects and may not be suitable for use at high levels of exposure (29, 42). Conversely, some biomarkers are present at concentrations too low to be reliably detected at low levels of exposure. For example, some biomarkers of alcohol abuse were not appropriate to evaluate low to moderate levels of alcohol consumption (43).

Specificity is another essential characteristic of biomarkers. Some biomarkers can be highly specific for a particular food (Table 1). Proline betaine and lycopene are well-established biomarkers for citrus fruit and tomato products, respectively (44, 45). Other biomarkers may be common to several foods or characteristic of an entire food group. Vitamin C and a number of carotenoids and flavonoids are common to many fruit and vegetables. Vitamin C or the sum of carotenoids or flavonoids have been used as generic biomarkers for fruit and vegetable intake (18, 45).

Single biomarker or combinations of biomarkers

Traditionally, single biomarkers have been used to characterize complex dietary exposures such as consumption of a whole food group or intake of a group of compounds with related biological activities. Two examples show the limits of such global assays. Vitamin C used as a biomarker for fruit and vegetable intake is present in a large number of fruit and vegetables, but its content varies widely according to species, varieties, and food-processing methods. It is also widely used as an additive and dietary supplement. The Folin assay, commonly used to estimate total polyphenols in foods (46), has also been applied to urine samples to compare polyphenol intake (47), but such use may be inappropriate because of the presence of interfering reducing metabolites in such complex biological matrices (46).

In contrast to these global assays, analytic approaches based on the estimation of combinations of dietary constituents may provide more accurate measurements of dietary exposure. The ratios of 2 alkylresorcinols characteristic of whole-grain wheat or rye were found to be good indicators of the relative consumption of these cereals (20, 48). However, there are very few such examples in which combinations of biomarkers were used to improve the specificity of dietary exposure measurements. Metabolomics constitutes a comprehensive approach to identify new panels of biomarkers that are specific or common to particular foods or food groups, as shown recently for citrus fruit (49). This should greatly improve the assessment of exposure to classes of food bioactive compounds, food groups, or dietary patterns.

THE FOOD METABOLOME IN THE OMICS ERA

Metabolomics can be described as the application of high-throughput analytic chemistry technologies [liquid chromatography–mass spectrometry (LC-MS)[4], nuclear magnetic resonance

[4] Abbreviations used: dbNP, Nutritional Phenotype Database; ECMDB, E. coli Metabolome Database; FDR, false discovery rate; FooDB, Food Component Database; GC-MS, gas chromatography–mass spectrometry; HMDB, Human Metabolome Database; LC-MS, liquid chromatography–mass spectrometry; MS, mass spectrometry; MSI, Metabolomics Standards Initiative; MWAS, metabolome-wide association study; NMR, nuclear magnetic resonance spectroscopy; PCA, principal components analysis; PLS-DA, partial least-squares discriminant analysis; TMAO, trimethylamine oxide-*N*-oxide; YMDB, Yeast Metabolome Database.

TABLE 1
Biomarkers used as surrogate indicators of consumption of foods and food groups for which significant ($r > 0.3$) correlations have been reported[1]

Food category and food	Biomarkers
Fruit	
Apple	Kaempferol, isorhamnetin, *m*-coumaric acid, phloretin
Orange	Caffeic acid, hesperetin, proline betaine
Grapefruit	Naringenin
Citrus fruit	Ascorbic acid, β-cryptoxanthin, hesperetin, naringenin, proline betaine, vitamin A, zeaxanthin
Fruit (total)	4-*O*-Methylgallic acid, β-cryptoxanthin, carotenoids (mix), flavonoids (mix), gallic acid, hesperetin, isorhamnetin, kaempferol, lutein, lycopene, naringenin, phloretin, vitamin A, vitamin C, zeaxanthin
Vegetables	
Carrot	α-Carotene
Tomato	Carotenoids (mix), lycopene, lutein
Vegetables, leafy	Ascorbic acid, β-carotene, carotenoids (mix)
Vegetables, root	Ascorbic acid, α-carotene, β-carotene
Vegetables (total)	Ascorbic acid, α-carotene, β-carotene, β-cryptoxanthin, carotenoids (mix), enterolactone, lutein, lycopene
Fruit and vegetables (total)	α-Carotene, apigenin, ascorbic acid, β-carotene, β-cryptoxanthin, carotenoids (mix), eriodictyol, flavonoids (mix), hesperetin, hippuric acid, lutein, lycopene, naringenin, phloretin, phytoene, zeaxanthin
Cereal products	
Whole-grain rye	5-Heptadecylresorcinol, 5-pentacosylresorcinol, 5-tricosylresorcinol
Whole-grain wheat	5-Heneicosylresorcinol, 5-tricosylresorcinol, alkylresorcinols (mix)
Whole-grain cereals (total)	5-Heneicosylresorcinol, 3,5-dihydroxybenzoic acid, 3-(3,5-dihydroxyphenyl)-1-propanoic acid, 5-pentacosylresorcinol, 5-tricosylresorcinol, alkylresorcinols (mix)
Seeds	
Soy products	Daidzein, genistein, isoflavones (mix), *O*-desmethylangolensin
Meats	
Meat	1-Hydroxypyrene glucuronide, 1-methylhistidine
Meat, beef	Pentadecylic acid
Animal products (total)	1-Methylhistidine, 3-methylhistidine, margaric acid, pentadecylic acid, phytanic acid
Dairy products	
Milk, dairy products	Iodine, margaric acid, pentadecylic acid, phytanic acid
Fish	
Fatty	DHA, EPA, long-chain ω-3 PUFAs, polychlorinated biphenyl toxic equivalents, pentachlorodibenzofuran, polychlorinated biphenyl 126, polychlorinated biphenyl 153, ω-3 PUFAs
Lean	Long-chain ω-3 PUFAs
Beverages (nonalcoholic)	
Tea	4-*O*-Methylgallic acid, gallic acid, kaempferol
Coffee	Chlorogenic acid
Beverages (alcoholic)	
Wine	4-*O*-Methylgallic acid, caffeic acid, gallic acid, resveratrol metabolites
Beverages (alcoholic) (total)	5-Hydroxytryptophol/5-hydroxyindole-3-acetic acid, carbohydrate-deficient transferrin, ethyl glucuronide, γ-glutamyltransferase, aspartate aminotransferase, alanine aminotransferase

[1] Data were extracted from the Exposome-Explorer database (V Neveu, DS Wishart, and A Scalbert, unpublished data, 2014).

spectroscopy (NMR), gas chromatography–mass spectrometry (GC-MS)] directed at characterizing the metabolome (ie, the small molecules associated with metabolism). Its development follows that of genomics, transcriptomics, and proteomics. Although not as rapid in development or as high-throughput as its omics cousins, metabolomics led a sea change in how small molecules could and should be analyzed. Rather than being limited to measuring only one or a few compounds at a time, new metabolomic technologies allowed researchers to measure hundreds or even thousands of metabolites at a time. This newly found capacity to measure so many chemicals at once led to a number of metabolomic projects, all launched in the mid-2000s, aimed at identifying the metabolomes of microbes (50), plants (51), and humans (52–54). These projects typically used LC-MS, GC-MS, NMR, or a combination of all 3 techniques to identify and/or quantify as many metabolites as possible in cells, tissues, and biofluids of the organisms of interest. These comprehensive metabolomic studies were also complemented by a number of much more specific metabolomic studies aimed at characterizing the metabolic responses of humans to the intake of various foods or food constituents such as soy (55), citrus fruit (44), nuts (56), meats (57), and tea (58).

The food metabolome as part of the human metabolome

It was through these early metabolome studies that scientists realized that the human metabolome was not as small or as simple as first imagined. In particular, noticeable differences in human metabolomes could be detected that appeared to depend strongly on diet, sex, health status, genetics, kinetics, physiology, and age—with diet being most important (59–62). This dietary dependence was not unexpected, but it was not anticipated to be so complicated. Unlike laboratory animals, humans are free-living omnivores who, in fact, eat other metabolomes. Furthermore, humans are exposed to a huge variety of "chemical environments" associated with the various foods we consume. Thus, the human metabolome is not just a single entity but consists of several components (Figure 1), including the following: *1*) the

endogenous metabolome (consisting of chemicals needed for, or excreted from, cellular metabolism), *2*) the food metabolome (consisting of essential and nonessential chemicals derived from foods after digestion and subsequent metabolism by the tissues and the microbiota), *3*) other xenobiotics derived from drugs, and *4*) xenobiotics derived from environmental or workplace chemicals.

The exact size and composition of these different human metabolomes are difficult to ascertain. Minimally, the human metabolome contains 50,000 different detectable compounds (9, 63), but as instrument sensitivity and separation technologies improve, this number is expected to increase. Up to 200,000 different metabolites are estimated to occur in the plant kingdom, and combinations of several hundreds of secondary metabolites generally characterize each edible plant (6, 64, 65). Furthermore, the composition often depends on the body compartment, tissue, or biofluid to which one refers. For instance, many food or drug constituents that might be found in the mouth or stomach are chemically identical to the compounds isolated from the intact food or drug. On the other hand, food constituents found in blood, urine, or other excreta are often metabolically transformed in the liver, kidney, or intestine to metabolites that are very different from the parent compound. This adds greatly to the diversity of the food metabolome. However, in some cases, the parent compounds are broken down to such an extent that their end products are actually identical to chemicals that the body produces naturally. The importance of the gut microbiota in contributing metabolites to the human metabolome has also recently emerged (50, 66). Some microbial metabolites, typically vitamins, certain essential amino acids, and a few fatty acids, are specific microbial metabolites ($\sim$100 compounds in total are known at this time). However, a large majority of the metabolites produced by the gut microbiota are derived from the biotransformation of both the endogenous metabolome and the food metabolome and are therefore an integral part of these 2 metabolomes. These microbial metabolites include short-chain fatty acids, secondary bile acids, protein and amino acid metabolites, as well as plant polyphenol metabolites (67).

Metabolism of food constituents

Knowledge of the metabolism of food constituents is critical to understanding the origin of the biotransformed fraction of the food metabolome. It is also essential if we wish to use food metabolites as nutritional biomarkers or as a means to monitor food consumption. In this regard, it is useful to review how food chemicals can be metabolized. Food constituents can be metabolized in 3 different ways: *1*) they can be digested in the mouth, stomach, and small intestine into simple nutrients that can be absorbed through the gut barrier; *2*) they can be further transformed by host tissues, especially the liver and kidney; or *3*) they can be processed by the gut microbiota in the large intestine.

The first category of food constituents are intermediary metabolites formed by digestion of lipids, polysaccharides, and proteins. Most of these compounds are common to all living organisms and identical to human endogenous metabolites. They cannot generally be used as dietary biomarkers because of their common identity and the impossibility to trace their dietary origin. The possible exceptions are the essential amino acids, essential fatty acids along with most vitamins, and minerals, which cannot be produced by humans and must originate from external dietary sources.

The second way that food constituents can be metabolized is through transformation by host tissues. Food compounds that are not useful for basic metabolism or that do not correspond to familiar endogenous metabolites are treated as "foreign" or as xenobiotics. Examples of exogenous food constituents include polyphenols, alkaloids, carotenoids, chlorophylls, artificial colors, artificial flavors, natural volatiles for flavoring/aroma, and Maillard reaction products formed during cooking. The human body maintains a complex defense system consisting of dozens of enzymes and membrane transporters to recognize these foreign and potentially toxic chemicals and to neutralize them by rapid biotransformation and/or elimination. Classically, the biotransformation process consists of 2 types of chemical reactions, phase I and phase II transformations, both of which occur primarily in the liver, kidney, and intestine. Phase I transformations typically involve oxidation of compounds via cytochrome P450 enzymes as well as hydrolysis by various dehydrogenases, esterases, and amidases. On the other hand, phase II transformations consist of chemical modifications such as methylation (by methyltransferases), sulfation (by sulfotransferases), acetylation (by *N*-acetyltransferases), glucuronidation (by UDP-glucuronyltransferases), and amino acid conjugation (by glutathione or glycyl transferases). A recent meta-analysis (68) of the metabolic fate of >1000 xenobiotics showed that cytochrome P450 catalyzed oxidations (40%) and UDP-glucuronosyltransferase glucuronidations (14%) were the most common followed by reactions involving dehydrogenases (8%), hydrolases (7%), glutathione-*S*-transferases (6%), and sulfatases (5%). In fact, there are >300 different empirical rules that allow one to predict the fate of metabolites on the basis of their chemical structure (69). Many of the metabolites derived from the biotransformation of food components have not been well characterized. For polyphenols, >230 phase I/II metabolites have been identified and associated with the consumption of specific polyphenol-containing foods (70). The yield of phase I/II reactions are often very high (68, 71), and host-transformed metabolites retain many of the features of their parent compounds. Consequently, these exogenously derived metabolites can be quite useful as specific food biomarkers.

The third way that food metabolites may be transformed is through microbial metabolism. Microbes have a very different set of enzymes from mammals, and given that there are >1000 different species of microbes in the human gut (72) there is an enormous diversity of enzymatic processes that act on food-derived compounds. The gut microbiota is particularly adept at processing polyphenols to phenolic breakdown products. For instance, depending on the predominant microbiota, polyphenols can be transformed by ring cleavage to a variety of aromatic compounds such as benzoate and various derivatives of hydroxyphenylacetic and hydroxypropionic acids. These phenolic acids can be further conjugated to glycine as in hippurate. The gut microbiota also processes indigestible carbohydrates through a variety of fermentative pathways yielding short-chain fatty acids such as butyric acid and propionic acid. Certain microbial metabolites can be useful as food biomarkers, although there is a complex relation between the food source, the predominant gut microbial species, and the resulting food

metabolites (73). Consequently, weaker correlations with intakes of foods or of their constituents were observed for microbial metabolites when compared with untransformed food compounds and host-transformed metabolites (41). This is most probably a result of the large variability of the microbiota across subjects (74). As a result, microbial metabolites should be treated with some caution when used as food biomarkers.

Food metabolome and metabolite databases

Given the complexity of food constituents, the diversity of known food metabolites, and the rapidly growing number of studies on the food metabolome, it is becoming clear that well-curated databases are of utmost importance to keep track of this information. These "omics era" databases are being developed to help researchers understand the origins and fate of many food metabolites (**Table 2**). Some recent examples include the Human Metabolome Database (HMDB) (9), the E. coli Metabolome Database (ECMDB) (66), the Yeast Metabolome Database (YMDB) (75), Food Component Database (FooDB) (76), Phenol-Explorer (70), and PhytoHub (77). HMDB is an online database of all known and presumptive human metabolites. This rapidly growing database currently contains >40,000 metabolites including endogenous, microbial, biotransformed, and exogenous/xenobiotic compounds. ECMDB is another online database consisting of 2750 metabolites known to be produced by *Escherichia coli*. This resource provides a representative estimate of the microbial metabolome that exists within the human gut. YMDB is a database consisting of 1730 metabolites known to be produced by *Saccharomyces cerevisae*. Given the number of food products (wine, beer, bread) produced by yeast fermentation and given that yeast also lives in the human gut, this database can also provide some useful data with regard to food metabolites and their possible origins and fate. FooDB is a database of >28,000 food constituents, including artificial food additives. Much of the chemical data in FooDB is now in HMDB, but FooDB provides additional information about food sources and food concentrations that is not in the HMDB. PhytoHub is an online database dedicated to the phytochemicals present in plant foods (~1000 compounds), their known human metabolites reported in the literature, and other potential metabolites predicted with in silico expert systems. Phenol-Explorer is an online database providing detailed information on dietary polyphenols and polyphenol metabolites. These food-focused resources are particularly detailed and provide substantially more in-depth information and reference material than what is available in the HMDB, YMDB, and ECMDB resources. Entries in each of these databases mentioned here are linked to other online resources such as PubMed, PubChem, Kyoto Encyclopedia of Genes and Genomes, Chemical Entities of Biological Interest, ChemSpider, and other widely used chemical resources. The establishment of these database resources along with the increasingly widespread use of metabolomics in nutrient analysis has now moved the field of food and nutrition science firmly into the modern "omics" era.

METABOLOMICS AND DISCOVERY OF NOVEL DIETARY BIOMARKERS

Study design

As noted previously, metabolomics has emerged as a key tool in the search for novel biomarkers of dietary intake. To date, the methods used for biomarker discovery can be divided into 2 main categories: hypothesis-driven and data-driven. In both cases, metabolomics-based approaches can be applied. In the hypothesis-driven approach, prior knowledge about the biomarker or a series of biomarker candidates is available from food composition databases such as FooDB (78) and methods are developed to measure the candidate biomarkers. So far, this approach has essentially been applied to specific families of food constituents such as fatty acids or carotenoids (45, 79).

In the data-driven approach, there is no prior knowledge of the biomarker and a large number of metabolites are measured, with the main limitation being the capacity of the analytic instrument to detect them. This approach has been used to discover novel biomarkers for a number of foods, nutrients, or diets (**Table 3**). The samples to be analyzed can be obtained from *1*) controlled dietary interventions or *2*) cross-sectional studies.

In controlled dietary interventions, subjects consume the food items of interest in a single meal (acute study) or in repeated meals over a given period of time (ranging from a few days to up to 6 mo; short- to medium-term study). In acute studies, biofluids are collected postprandially over a time period of up to 24 h after consumption of the food of interest. Ideally, any biomarker identified in these acute studies must be validated with an intervention study to ensure there is a dose response, which would render the biomarker suitable for use over a range of intakes. In short-term interventions, biofluids are collected at the end of the intervention period and compared in subjects consuming either the test food or a control food. Biofluids can also be collected before and after consumption of the test food. A limitation of these intervention studies is the fact that the biomarkers identified

TABLE 2
Metabolite databases related to the food metabolome and accessible online[1]

Database	Metabolites	No. of metabolites	Website address	Reference
HMDB	Endogenous, microbial, biotransformed, and exogenous/xenobiotic compounds identified in humans	>40,000	www.hmdb.ca	(9)
ECMDB	*Escherichia coli* metabolites	2750	www.ecmdb.ca	(66)
YMDB	*Saccharomyces cerevisiae* metabolites	1730	www.ymdb.ca	(75)
FooDB	Food constituents and food additives	28,000	www.foodb.ca	(76)
Phenol-Explorer	Dietary polyphenols and their metabolites	502	www.phenol-explorer.eu	(70)
PhytoHub	Dietary phytochemicals and their metabolites	1500	www.phytohub.eu	(77)

[1] ECMDB, E. coli Metabolome Database; FooDB, Food Component Database; HMDB, Human Metabolome Database; YMDB, Yeast Metabolome Database.

may not be sufficiently specific for the test food in population studies, because regular diets may include other foods containing precursors of the same biomarkers. For instance, in a cross-sectional analysis of a whole-diet intervention study it was only possible to verify 23% of potential biomarkers observed in previous-meal studies (81).

Cross-sectional studies can therefore play an important role in biomarker discovery. Low and high consumers are selected from food intake data collected by using food-frequency questionnaires, food diaries, or other dietary assessment tools. Comparison of these groups can lead to the identification of biomarkers that are reflective of habitual intake, provided that these biomarkers have a sufficient half-life in the organism or that the foods are regularly consumed. Although these and other studies have shown the potential of cross-sectional studies, care needs to be taken because many of the foods consumed are highly correlated and there is a risk of identifying biomarkers that are not specific to the particular food of interest unless their identity and specific occurrence in the considered foods are established. Notwithstanding, cross-sectional studies are excellent resources that are currently underused for dietary biomarker discovery.

Novel dietary biomarkers identified through a metabolomic approach

An extensive list of potential dietary biomarkers discovered by metabolomics is presented in Table 3. Markers associated with the consumption of foods, nutrients, or diets have been identified. Successful studies include the identification of proline betaine as a marker of citrus intake (49, 80). This marker was first identified in small-scale acute feeding studies and validated in free-living subjects in 2 independent studies (44, 80). It was confirmed in a cross-sectional study that used untargeted metabolomics (49) and played an important role in discriminating noncompliant individuals in a dietary pattern study of Nordic compared with habitual diets (106). In these same studies, screening of urinary profiles for predicted metabolites of citrus fruit also led to the identification of some terpenoids and flavonoids as biomarkers of citrus food intake as well as of intake of citrus-flavored sweets. This shows well the importance of previous knowledge on food composition and on metabolism of food constituents for annotating unknown discriminating ions in untargeted metabolomic studies.

Trimethylamine oxide-*N*-oxide (TMAO) was found to be a putative biomarker for meat intake or for meat-containing diets in several studies (102–104), but it has also been reported as a biomarker of fish intake by other authors (82, 107) and shown to be more responsive to intake of fish than meat (85). Several dietary precursors of TMAO such as choline or carnitine have been described (108) and care should be paid when interpreting variations in TMAO concentrations in populations.

The state of validation of biomarkers listed in Table 3 varies widely. Proline betaine is a good example of a well-validated citrus fruit biomarker. Other biomarkers, particularly those identified in controlled intervention studies, may prove to be less robust in populations because of the possible existence of a variety of precursors as seen for TMAO, or the occurrence of the same precursor in various foods. Food-derived biomarkers such as caffeic acid sulfate or methylepicatechin sulfate, which were found to discriminate consumers of raspberries (82), may not be that useful in epidemiologic studies because both their parent metabolites (caffeic acid and epicatechin) have been described in a variety of foods of plant origin (70).

For this reason, it may be particularly advisable to look for characteristic dietary biomarkers directly in cross-sectional studies. However, the chances to identify robust biomarkers will rely both on the sensitivity of the analytic equipment used and on the quality of the dietary data against which metabolic profiles are correlated. Both 24-h dietary recalls and food-frequency questionnaires have been used, and new biomarkers for citrus fruit intake or coffee were successfully identified (49, 88) (Table 3). The use of food-frequency questionnaires may directly lead to the identification of biomarkers of habitual dietary exposure, but the lower accuracy and lower number of foods documented may limit their value for such discovery studies (105).

With the exception of 2 studies on dietary fiber and milk protein diet, all discovery studies were conducted on urine samples as opposed to blood samples (Table 3). The reason for this is partly technical because of the higher concentrations of food-derived metabolites in urine as compared with blood and because of the lack of interfering proteins. This contrasts with the preferred use of blood biospecimens to measure biomarkers of nutritional status in epidemiologic studies. More metabolomic studies using blood samples should be carried out because of the more common availability of plasma or serum samples in biobanks. Also, lipophilic biomarkers, which may be more stable over time (*see* Pharmacokinetics and reliability of dietary biomarkers section), are more likely to be found in blood. Regression analyses of the concentrations of 363 metabolites in plasma with a number of dietary variables measured with a food-frequency questionnaire showed the highest correlations with phospholipid concentrations (109). Furthermore, chain length and degree of saturation of fatty acids in glycerophosphatidylcholines were associated with intake of specific foods or nutrients such as fish and dietary fiber.

It is important to point out that the identities of many of the proposed biomarkers in Table 3 (marked with an asterisk) have not been fully validated with proper chemical standards because these standards are often not commercially available. In addition, no standard yet exists to report chemical identification of biomarkers in metabolomic studies (110). For this reason, it is often difficult to evaluate the degree of confidence in biomarker identification.

Analysis of the food metabolome

Analyzing the food metabolome is a particularly challenging task for 3 reasons. First, it comprises a much greater chemical diversity than any other part of the metabolome (*see* Food metabolome and metabolite databases section). A second feature of the food metabolome is the huge range of concentrations, from picomolar or nanomolar concentrations for some contaminants or phytochemical metabolites to millimolar concentrations for nutrients such as sugars. Third, many components of the food metabolome are unknown. Indeed, the metabolism for a large proportion of nonnutrients in humans has never been studied and the chemical structures of their circulating metabolites have not been identified. Until recently, the food metabolome was typically analyzed through targeted methods optimized for specific compounds or families of nutrients or nonnutrients, such as

TABLE 3
Tentative dietary biomarkers identified through untargeted metabolomic approaches in human dietary intervention studies and cross-sectional studies[1]

Dietary factor and study type	No. of subjects	Comparison	Dietary assessment tool	Biospecimen	Analytic technique	Biomarker	Reference
Fruit, fruit juices							
Mixed-fruit meal							
AI	8	Consumers/ control	NA	U (spot)	NMR	Proline betaine	(80)
Citrus fruit							
CS	499	Consumers/ nonconsumers	24-h dietary record	U (24-h)	NMR	Proline betaine	(80)
CS	12	H/M/L	FFQ	U (fasting)	FIE-FTICR-MS	Proline betaine, 4-hydroxyproline betaine	(44)
Orange juice							
AI	4	Consumers/ control	NA	U (kinetics)	LC-Q-Tof, LTQ-Orbitrap	Proline betaine, limonene-8,9-diol-glucuronide,* nootkatone-13, 14-diol-glucuronide,* hesperetin-3′- glucuronide, hydroxyproline betaine, *N*-methyltyramine-sulfate,* naringenin-7-*O*-glucuronide	(49)
SMTI	12	Consumers/ control	NA	U (24-h)	LC-Q-Tof, LTQ-Orbitrap		
Citrus fruit							
CS	80	H/L	FFQ and 24-h dietary record	U (spot)	LC-Q-Tof, LTQ-Orbitrap		
CS	107	Consumers/ nonconsumers	24-h dietary record	U (24-h)	LC-Q-Tof	Proline betaine, hesperetin-3-glucuronide*	(81)
Raspberries							
SMTI	24	Consumers/ control	NA	U (kinetics)	FIE-FTICR-MS, GC-Tof	Caffeic acid-sulfate, methylepicatechin-sulfate	(82)
Strawberries							
CS	107	Consumers/ nonconsumers	24-h dietary record	U (24-h)	LC-Q-Tof	2,5-Dimethyl-4-methoxy-3(2H)-furanone-sulfate*	(81)
Vegetables							
Vegetables							
CS	160	H/M/L	Food diary	U (fasting)	NMR	Phenylacetylglutamine	(83)
Broccoli							
SMTI	24	Consumers/ control	NA	U (kinetics)	FIE-FTICR-MS	Tetronic acid,* xylonate/lyxonate,* threitol/erythritol*	(82)
Cruciferous vegetables							
SMTI	20	Before/after	NA	U (kinetics)	NMR	*S*-Methyl-L-cysteine sulfoxide	(84)
AI	17	Consumers/ control	NA	U (kinetics)	LC-Q-Tof	Sulforaphane *N*-acetylcysteine, *N*-acetyl-(*N*′-benzylthiocarbamoyl)cysteine, sulforaphane *N*-cysteine,* *N*-acetyl-*S*-(*N*-3-methylthiopropyl)cysteine,* *N*-acetyl-*S*-(*N*-allylthiocarbamoyl)cysteine,* iberin *N*-acetyl-cysteine,* 4-iminopentylisothiocyanate,* erucin *N*-acetylcysteine*	(85)
Red cabbage							
CS	107	Consumers/ nonconsumers	24-h dietary record	U (24-h)	LC-Q-Tof	3-Hydroxy-3-(methylsulfinyl)propanoic acid,* 3-hydroxyhippuric acid-sulfate,* 3-hydroxyhippuric acid,* iberin *N*-acetyl-cysteine,* *N*-acetyl-*S*-(*N*-3-methylthiopropyl)cysteine,* *N*-acetyl-*S*-(*N*-allylthiocarbamoyl)cysteine,* sulforaphane *N*-acetylcysteine*	(81)
Beetroot							

(Continued)

TABLE 3 (*Continued*)

Dietary factor and study type	No. of subjects	Comparison	Dietary assessment tool	Biospecimen	Analytic technique	Biomarker	Reference
CS	107	Consumers/ nonconsumers	24-h dietary record	U (24-h)	LC-Q-Tof	4-Ethyl-5-aminopyrocatechol sulfate,* 4-ethyl-5-methylaminopyrocatechol-sulfate,* 4-ethylpyridine-2-carboxylic acid glycine conjugate	(81)
Cereals							
Whole-grain rye bread							
SMTI	20	Consumers/control	NA	U (24-h)	LC-Q-Tof	3-(3,5-Dihydroxyphenyl)-1-propanoic acid-sulfate* and -glucuronide,* enterolactone- glucuronide,* azelaic acid,* 2-aminophenol-sulfate,* 2,4-dihydroxy-1,4-benzoxazin-3-one,* 2-aminophenol-sulfate,* 2-4-dihydroxy-1,4-benzoxazin-3-one-sulfate,* indolylacryloylglycine,* ferulic acid-sulfate,* 3,5-dihydroxyphenylethanol-sulfate,* 3,5-dihydroxycinnamic acid-sulfate*	(86)
Meat and fish							
Red meat							
CS	160	H/M/L	Food diary	U (fasting)	NMR	*O*-Acetylcarnitines	(83)
Salmon							
SMTI	24	Consumers/control	NA	U (kinetics)	FIE-FTICR-MS	Anserine, methylhistidine, trimethylamine-*N*-oxide	(82)
Oily fish							
CS	68	H/M/L	FFQ	U (spot, 24-h, fasting)	FIE-FTICR-MS	Methylhistidine	(87)
Beverages							
Coffee							
CS	18	Consumers/ nonconsumers		U (fasting)	LC-Q-Tof	*N*-Methylpyridinium, trigonelline	(88)
AI	9	Before/after	NA	U (kinetics)	LC-Q-Tof	*N*-Methylpyridinium, trigonelline	(88)
CS	68	H/M/L	FFQ	U (spot, 24-h, fasting)	FIE-FTICR-MS	Dihydrocaffeic acid	(87)
Chamomile tea							
SMTI	14	Before/after	NA	U (spot)	NMR	Hippuric acid*	(89)
Black tea							
AI	3	Before/after	NA	U (24-h)	NMR	Hippuric acid,* gallic acid, 1,3-dihydroxyphenyl-2-*O*-sulfate*	(90)
Tea (black and green)							
STI	17	Consumers/control	NA	U (24-h)	NMR	Hippuric acid.* 1,3-dihydrophenyl-2-*O*-sulfate*	(58)
Green tea							
AI	8	Before/after	NA	U (kinetics)	NMR	Hippuric acid*	(91)
Black tea							
AI	20	Consumers/control	NA	U (kinetics)	NMR	Hippuric acid,* 4-hydroxyhippuric acid,* 1,3-dihydrophenyl-2-*O*-sulfate,* allic acid, 4-*O*-methylgallic acid*	(92)

(Continued)

TABLE 3 (*Continued*)

Dietary factor and study type	No. of subjects	Comparison	Dietary assessment tool	Biospecimen	Analytic technique	Biomarker	Reference
Mixed red wine/grape juice extracts							
SMTI	35	Consumers/control	NA	U (24-h)	GC-MS, LC-MS/MS	Hippuric acid,* 3-hydroxyhippuric acid,* 4-hydroxyhippuric acid,* 4-hydroxybenzoic acid,* 1,2,3-trihydroxybenzene,* vanillic acid,* isovanillic acid,* syringic acid,* 3-hydroxyphenylacetic acid,* 4-hydroxymandelic acid,* vanilmandelic acid,* ferulic acid,* 3-hydroxyphenylpropionic acid,* 3,4-dihydroxyphenylpropionic acid,* 3-(3-hydroxyphenyl)-3-hydroxypropionic acid,* catechol,* pyrogallol,* citrate,* betaine*	(93, 94)
Wine							
SMTI	61	Consumers/control	NA	U (24-h)	NMR	Tartrate,* 4-hydroxyphenylacetate,* mannitol,* ethanol*	(95)
Other foods							
Cocoa powder							
AI	10	Consumers/control	NA	U (kinetics)	LC-Q-Tof	Vanilloylglycine,* 6-amino-5-(*N*-methylformylamino)-1-methyluracil,* 3-methyluric acid,* 7-methyluric acid,* 3-methylxanthine,* 7-methylxanthine,* dimethyluric acid,* theobromine, caffeine, trigonelline,* hydroxynicotinic acid,* tyrosine, 3,5-diethyl-2-methylpyrazine,* hydroxyacetophenone,* diketopiperazines,* epicatechin-sulfate,* *O*-methylepicatechin,* vanillic acid,* phenylvaleric acid* and phenylvalerolactone* derivatives, furoylglycine,* xanthurenic acid*	(96)
SMTI	20	Consumers/control and before/after	NA	U (24-h)	LC-Q-Tof	Hydroxynicotinic acid,* 6-amino-5-(*N*-methylformylamino)-1-methyluracil,* 7- and 3-methyluric acid,* 7- and 3-methylxanthine,* 3,7-dimethylruric acid,* cyclo(propylalanyl),* 3,5-diethyl-2-methylpyrazine,* theobromine,* vanillic acid-glucuronide* and -sulfate- glucuronide,* vanilloylglycine,* 4-hydroxy-5-(dihydroxyphenyl)-valeric acid-glucuronide* and -sulfate,* 3′-methoxy-4′-hydroxyphenylvalerolactone,* 4′-hydroxy-5-(hydroxymethoxyphenyl)valeric acid-glucuronide,* 5-(3′,4′-dihydroxyphenyl)-γ-valerolactone-glucuronide* and -sulfate* and -sulfate glucuronide,* (epi)catechin- glucuronide* and -sulfate glucuronide,* methyl-(epi)catechin-sulfate,* *N*-(4′-hydroxy-3′-methoxy-E-cinnamoyl)-L-aspartic acid,* *N*-(4′-hydroxycinnamoyl]-L-aspartic acid,* methoxyhydroxyphenylvalerolactone-glucuronide,* hydroxyphenylvalerolactone-glucuronide* and -sulfate,* 5-(hydroxymethoxyphenyl)valeric acid-sulfate,* 4-hydroxy-5-(phenyl)valeric acid-sulfate*	(97)
Chocolate (solid or drink)							
CS	107	Consumers/ nonconsumers	24-h dietary record	U (24-h)	LC-Q-Tof	6-Amino-5-(*N*-methylformylamino)-1-methyluracil,* theobromine, 7-methyluric acid	(81)

(*Continued*)

TABLE 3 (*Continued*)

Dietary factor and study type	No. of subjects	Comparison	Dietary assessment tool	Biospecimen	Analytic technique	Biomarker	Reference
Almond-skin extract							
AI	24	Before/after	NA	U (kinetics)	LC-Q-Tof	(Epi)catechin-sulfate,* *O*-methyl-(epi)catechin-sulfate,* naringenin-*O*-glucuronide,* 5-(hydroxyphenyl)-γ-valerolactone-glucuronide* and -sulfate,* 5-(dihydroxyphenyl)-γ-valerolactone- glucuronide,* -sulfate glucuronide* and -sulfate,* 5-(trihydroxyphenyl)-γ-valerolactone-glucuronide,* 5-(hydroxymethoxyphenyl)-γ-valerolactone-glucuronide* and sulfate,* 4-hydroxy-5-(dihydroxyphenyl)-valeric acid-glucuronide* and sulfate,* 4-hydroxy-5-(hydroxymethoxyphenyl)valeric acid-glucuronide,* 4-hydroxy-5-(methoxyphenyl)valeric acid-glucuronide,* 4-hydroxy-5-(hydroxyphenyl)valeric acid-glucuronide and -sulfate,* 4-hydroxy-5-(phenyl)valeric acid-sulfate,* 2-(dihydroxyphenyl)acetic acid-glucuronide,* -sulfate glucuronide* and -sulfate,* 2-(hydroxymethoxyphenyl)acetic acid-glucuronide,* 2-(hydroxyphenyl)acetic acid-sulfate,* 3-(hydroxyphenyl)propionic acid-glucuronide,* 3-(dihydroxyphenyl)propionic acid-sulfate,* vanillic acid-glucuronide,* hydroxyhippuric acid,* ferulic acid-glucuronide*	(98)
Nuts							
SMTI	42	Consumers/control	NA	U (24-h)	LC-Q-Tof, LTQ-Orbitrap	10-Hydroxydecene-4,6-diynoic acid-sulfate,* tridecadienoic/tridecynoic acid-glucuronide,* dodecanedioic acid,* 1,3-dihydroxyphenyl-2-*O*-sulfate,* *p*-coumaroyl alcohol-glucuronide* and -sulfate,* *N*-acetylserotonine-sulfate,* 5-hydroxyindoleacetic acid,* urolitin A-glucuronide, sulfate* and sulfate glucuronide*	(56)
Walnuts							
CS	107	Consumers/ nonconsumers	24-h dietary record	U (24-h)	LC-Q-Tof	5-Hydroxyindole-3-acetic acid	(81)
Nutrients							
Dietary fiber							
SMTI	77	H/L	Dietary record	U (24-h)	NMR	Hippuric acid*	(99)
SMTI	25	Consumers/control	NA	P (fasting)	LC-Q-Tof	2-Aminophenol-sulfate, 2,6-dihydroxybenzoic acid, hydroxynuategenin-glucuronide*	(100)
Whey protein isolate							
SMTI	12	Consumers/control	NA	P (sequential)	LC-Q-Tof	Tryptophan, phenylalanine, kynurenine, γ-Glu-Leu	(101)
Whey hydrolysate							
SMTI	12	Consumers/control	NA	P (sequential)	LC-Q-Tof	Methionine sulphoxide, cyclo(Pro-Thr), cyclo(Ala-Ile), cyclo(Phe-Val), β-Asp-Leu, pGlu-Pro,	(101)
Diets							
Omnivorous diet							
SMTI	12	Consumers/control	NA	U (24-h)	NMR	Taurine,* carnitine,* acetylcarnitine,* 1-methylhistidine,* 3-methylhistidine,* trimethylamine-*N*-oxide*	(102)
Vegetarian diet							
SMTI	12	Consumers/control	NA	U (24-h)	NMR	*p*-Hydroxyphenylacetate*	(102)
Meat protein diet							
SMTI	24	Before/after	NA	U (24-h)	NMR	Trimethylamine-*N*-oxide,* histidine*	(103)
Seafood							
AI	17	Consumers/control	NA	U (kinetics)	LC-Q-Tof	Trimethylamine-*N*-oxide	(85)

(*Continued*)

The American Journal of Clinical Nutrition

TABLE 3 (*Continued*)

Dietary factor and study type	No. of subjects	Comparison	Dietary assessment tool	Biospecimen	Analytic technique	Biomarker	Reference
Milk protein diet							
SMTI	24	Before/after	NA	S (fasting)	NMR	Short-chain fatty acids*	(103)
Omnivorous diet							
CS	161	Consumers/control	Questionnaire	U (fasting)	NMR	Trimethylamine-*N*-oxide,* dimethylamine,* phenylalanine,* methylhistidine*	(104)
Lactovegetarian diet							
CS	161	Consumers/control	Questionnaire	U (fasting)	NMR	Citrate*	(104)
Phytochemical-rich diet (citrus, cruciferous vegetables, soy)							
SMTI	10	Consumers/control	NA	U (spot)	LC-FTICR-MS	Sulforaphane,* proline betaine,* hippuric acid,* genistein,* daidzein,* equol,* glycitein,* *O*-desmethylangolensin,* enterolactone,* trigonelline*	(105)
CS	60	H/L	Dietary record	U (spot)	LC-FTICR-MS	Proline betaine*	(105)
Nordic diet							
SMTI	107	Consumers/control	24-h recall and supermarket records	U (fasting)	LC-Q-Tof	Trimethylamine-*N*-oxide, hydroquinone-glucuronide, hippuric acid, (2-oxo-2,3-dihydro-1H-indol-3-yl)acetic acid, 3,4,5,6-tetrahydrohippurate*	(106)

[1] *No standard was used to confirm the identity of the biomarker. AI, acute intervention; CS, cross-sectional; FFQ, food-frequency questionnaire; FIE, flow injection electrospray; FTICR, Fourier transform ion cyclotron resonance; GC, gas chromatography; H/L, high and low (intake); H/M/L, high, medium, and low (intake); LC, liquid chromatography; MS, mass spectrometry; NA, not applicable; NMR, nuclear magnetic resonance spectroscopy; P, plasma; Q, quadrupole; S, serum; SMTI, short- and medium-term intervention; Tof, time-of-flight; U, urine.

lipids, organic acids, sugars, flavonoids, or carotenoids. However, the combination of available targeted analysis methods is still far from covering the whole chemical space of the food metabolome. In principle, untargeted metabolomics provides a wider coverage and is likely to show the presence of new metabolites in biofluids and tissues.

As is the case for the other parts of the metabolome, mass spectrometry (MS) coupled with gas chromatography or liquid chromatography and NMR are currently the most widely used technologies for food metabolome analysis (Table 3). The advantages and disadvantages of these techniques have been extensively discussed elsewhere and are beyond the scope of the present review (111–113). Briefly, NMR is robust, nondestructive, and quantitative but has a relatively low sensitivity, which narrows its coverage of the food metabolome to predominant nutrients, sugars, and microbial metabolites present at millimolar to micromolar concentrations. MS is by far the most sensitive technique and the only method able to cover the nonnutrient metabolites of the food metabolome occurring at low concentrations in biological samples. GC-MS combined with chemical derivatization has been used to analyze constituents of the food metabolome such as phenolic acids or fatty acids (79, 93, 94). However, to date, most studies on the food metabolome have been performed by using high-resolution liquid chromatography-quadrupole-time-of-flight MS with electrospray ionization (Table 3). This technique has been successful in detecting compounds such as terpene metabolites, diketopiperazine metabolites, phenylvalerolactones, and benzoxazinoid metabolites, which are interesting candidate biomarkers of food intake that would not be easily detected in biofluids by NMR or GC-MS (49, 86, 96). No single chromatographic method is able to cover the wide range of polarity existing for the food metabolome compounds. Highly polar compounds may have to be analyzed by using hydrophilic interaction chromatography, whereas specific methods with atmospheric pressure chemical ionization may be developed for profiling apolar plasma metabolites. Direct flow injection-MS has also been used (82, 114), which offers the advantage of high-throughput analysis, as would be required for large-scale epidemiologic studies. However, ion suppression effects, due to inefficient ionization of certain ions in complex matrices and the inability to discriminate between isomers, limit the use of this approach.

The main current limitation of MS is the very challenging and burdensome task of the structural elucidation of the detected ions (*see* below). However, because of its sensitivity and breadth of coverage, LC-MS has certainly become the method of choice for untargeted analysis of the food metabolome. Rapid advances in technology have led to a new generation of much more efficient time-of-flight and single-stage Orbitrap (Exactive; Thermo Scientific) instruments, offering improved linearity, resolution, and mass accuracy, which will be critically important for the analysis of the food metabolome (115). As with any experimentally based analytic method, multiple variables can substantially affect the final data set. These include the mode of sample preparation, method of chromatography, mode of detection, and the choice of data reduction methods (116, 117). No standardized method exists yet, and the need for improved harmonization is certainly desirable for further progress on the food metabolome.

Achieving absolute quantification rather than relative quantification of food metabolome metabolites via untargeted methods

remains a continuing challenge. It is essentially impossible to use standards or isotopically labeled references to quantify the thousands of compounds in the food metabolome. New approaches are being developed with isotope labeling and multiple reaction monitoring–based profiling for families of compounds sharing distinctive chemical functionalities (118). Labeled reagents targeted at these functionalities or particular multiple reaction monitoring transitions could be used to specifically measure selected fractions of the food metabolome such as amines, phenols, glucuronides, or mercapturic acid derivatives. These advances may allow researchers to target larger areas of the food metabolome chemical space with the use of standardized quantitative methods.

Analysis of metabolomic data

The metabolic profile of raw data generated by the spectrometric analysis of biological samples can be analyzed in several steps (119, 120). These include data preprocessing, data alignment, data normalization, and signal correction followed by the analysis through various statistical methods. There are a number of different software tools available for these tasks; most vendors have their proprietary software but highly efficient freeware programs, Web servers, or add-on softwares exist. For NMR, an example is the Interval Correlation Optimized shift algorithm produced for Matlab (121), and for LC-MS data alignment freeware such as XCMS (122), MZmine (119, 123), and MetAlign (124) are widely used.

The preprocessing step is software dependent and typically includes data reduction methods such as centroiding of mass spectra or analog-to-digital conversion of NMR, infrared, or UV/visible spectra. Preprocessing also includes translation of data formats and data export. The next step is data alignment. It is crucial to align the different sample profiles, which do not match exactly because of small variations in retention times, masses, or chemical shifts. All available software tools differ in their peak picking algorithms. There is only a 50–70% overlap between the peaks detected by different packages from the same raw data set, even with similar settings (125). Additional markers may be observed by using additional softwares or simply by altering software settings. Another major difference between packages is the presence or absence of so-called gap filling, a routine to revisit the raw data for any peak that has not been detected in a sample when it was found in others. The lack of a gap-filling algorithm creates major problems for normalization and for statistical analysis. An ideal food intake marker would have a zero value in control samples from volunteers who did not consume the food; in this case, the gap-filling routine helps to estimate the background noise in the peak area.

The output from the peak detection and data alignment steps is typically a matrix of samples and features with the intensity as the values within the matrix structure. A feature here denotes any distinct peak in the data set, regardless of whether it represents a known, unknown, or even an artifact ion. In LC-MS profiling, the features are characterized by a retention time and a mass (m/z) value. Such a feature may be a compound's parent ion, but just as frequently it represents an adduct ion or a fragment from a compound. In NMR and in most other digitized spectral data the single features are part of spectral shapes that usually have local maxima and minima. For both kinds of data the fine structure of the data contains additional information that is useful for identifying compounds and structures in the samples and is therefore particularly important for characterizing the food metabolome.

Metabolic profiling data may be analyzed by using univariate or multivariate statistical methods. Statistical analysis of untargeted metabolomic data is often an initial step in the biomarker discovery process that should not be confused with hypothesis testing, because there is no a priori hypothesis. In dietary intervention studies with single foods, the contrast observed for a good biomarker can be large, sometimes even infinite, making it possible to work robustly with small sample sets and discriminate potential intake biomarkers from more subtle changes in endogenous metabolites (126). In cross-sectional studies this large contrast seldom applies, but approximate dose-response relations from food-frequency questionnaires may help in the identification of food intake biomarkers.

Multivariate analysis is most commonly used for explorative analysis of metabolic profiling data (127). As opposed to univariate analysis, multivariate analysis can be performed in an unsupervised manner (ie, without including information on group assignment for the analysis). This provides an objective assessment of the principal patterns in the data set (eg, intake or no intake of a specific food component or diet). Unsupervised analysis such as principal components analysis (PCA) should always be the starting point for explorative multivariate analysis to ascertain that there is an overall segregation into a food-related pattern. The features associated with any pattern can be shown by the loadings in a PCA plot; however, PCA is generally not well suited to identify the most prominent part of the pattern. Sparse PCA overcomes this limitation (128, 129). Clustering methods are also widely used for subdividing and ordering a data set into groups of data with a high degree of similarity. Hierarchical clustering generates a dendrogram in which neighboring samples share the greatest similarity and neighboring features are those most closely related. This provides useful biological information and unsupervised groupings of the data set (130).

Supervised multivariate analysis is commonly the next step in many data analysis methods but has a strong tendency to overfit the data. Even random data will usually segregate and show a "marker pattern" after supervised analysis (131). Careful validation with the use of techniques such as permutation testing and cross-validation is therefore always necessary. There are a large number of supervised methods (120, 127), with the most commonly used analysis for comparing 2 groups being partial least-squares discriminant analysis (PLS-DA) (132) or one of its several variants. In complex nutritional studies it may be useful to combine ANOVA separations of factors with PLS-DA (133, 134) or use multilevel PLS-DA to reduce the influence of interindividual variation (135). Some multivariate methods such as PLS are mainly used to fit the data to a continuous variable. This is useful to explore the relation of any features in the profiling data set with an external variable (eg, intake of a specific food based on a questionnaire or any biological outcome marker) (121). In addition, for these prediction models very careful validation is required and their global ability to predict a specific food intake has to be assessed in separate studies.

Univariate analysis is supervised—that is, a hypothesis regarding a difference between groups is implicit. Any marker identified by this approach should therefore also be independently validated in a separate study. For univariate analysis used in exploration of new food intake biomarkers it is important to set a reasonable threshold for false discovery rates

(FDRs) (136, 137). In explorative science there is no fixed rule for the acceptability level of the FDR, and any level from 5% to 50% may be useful, depending on further data analysis steps. If no additional data analysis is planned as, for instance, in metabolome-wide association studies (MWASs), the FDR should generally be selected in the lower end of this range. If the univariate data step is used for selection of features that will be analyzed further (eg, by multivatiate models), it may be more appropriate to use a higher FDR. In any case, the markers found must be validated in subsequent independent studies.

Overall, the field of metabolomics is rich with data analysis options, and the challenge in the future will be to optimally apply these to food metabolomic studies. Useful resources exist to help in selecting and using in the most rigorous way appropriate tools for data analysis in a particular project (138).

Metabolite identification

Metabolite identification in metabolomic studies relies on the comparison of generated spectra with those in curated metabolite databases. However, the vast majority of the food metabolome components are not yet represented in these databases, which makes the elucidation of their structure difficult. As previously described, the identification of candidate dietary biomarkers is complicated by the fact that the majority of food compounds are treated as xenobiotics by the human body in phase I and phase II reactions or undergo fermentation in the colon by the gut microbiota (139). Despite some increase in their availability over the past few years, these highly diverse metabolites are largely absent from most databases. One exception is Phenol-Explorer, which gives a comprehensive overview of the human and animal metabolites formed from polyphenols (70).

NMR spectroscopy and MS are the 2 essential tools for elucidation of the structure of unknown metabolites in metabolomic studies (140, 141). Metabolites such as *S*-methyl-L-cysteine sulfoxide or proline betaine as biomarkers of cruciferous vegetables or citrus fruit, respectively, could be identified in NMR studies on the basis of their characteristic chemical shifts (Table 3) (80, 84). More markers of food intake have been identified in MS-based metabolomic studies on the basis of their accurate mass and mass fragmentation spectra (142–144).

A number of commercial and "in house" software tools have been developed recently and used to recognize and identify fragments and adducts derived from one food metabolite (97, 114, 145). These tools are particularly useful to identify phase II conjugates, common constituents of the food metabolome, which show characteristic neutral losses (eg, 79.957 amu for sulfate conjugates and 176.032 amu for glucuronides) (44, 114, 140). Customized in-house databases on the most likely phase I and phase II metabolites have also been developed based on in silico prediction with expert systems such as Meteor (49, 146–148). An important challenge for the future will be the development of a coordinated international effort to extend existing and develop novel software tools and databases allowing the more "intelligent" prediction of the metabolic fate of food constituents.

Spectral databases for the food metabolome

Despite the many initiatives to make spectral data sets available to the scientific community, the publicly accessible existing spectral data are still scattered over numerous Web-based (searchable) databases, printed tables in scientific journals, Excel files in supporting information, and scientific books (140, 149, 150). As described earlier, the major online chemical resources (typified by PubChem, Kyoto Encyclopedia of Genes and Genomes, and ChemSpider) contain limited information on human metabolites derived from food compounds. Although they are not specific to the food metabolome, these resources are useful for metabolite identification because fragmentation data or NMR signals of known metabolites can be compared with the unknown query to gain structural information. The most comprehensive and best-curated chemical (commercial) database is currently SciFinder, which includes many food metabolome compounds collected from the literature (151). Recently, a large number of food compounds have been added to the HMDB, which makes searches on the basis of their accurate masses possible (9); however, to date it contains few mass fragmentation spectra useful for food metabolite identification.

The robust and reproducible fragmentation patterns and retention times of volatile metabolites in GC-MS have successfully been used to set up metabolomic workflows that search for possible candidate metabolites in the National Institutes of Standards and Technology library or in-house libraries (152). Recently, similar approaches have been proposed for LC-MS– and NMR-based metabolomic data sets. Spectral databases contain fragmentation spectra obtained in different experimental conditions (eg, several collision energies and different mass spectrometers) to facilitate direct comparison with experimental data (141). Also, the number of metabolite spectra in Chemspider and HMDB is increasing. Even though not specific to the food metabolome, these resources are particularly useful for metabolite identification because fragmentation data or NMR signals of known metabolites can give structural hints for the unknown query.

Software tools for annotation of the food metabolome

Software tools such as MetFrag, MyCompoundID, MetiTree, and Mass Frontier can handle metabolite fragmentation data and permit library searches for potential candidates using in silico fragmentation predictions of metabolites or comparisons to previously fragmented metabolites or standards (146, 153–155). MetFusion combines knowledge from spectral databases such as MassBank with the multitude of candidates generated by fragmenters such as MetFrag (156).

Software tools have also been developed that integrate metabolite annotation directly within the processing pipeline of LC-MS data (157). For example, CAMERA is a pipeline for the annotation and analysis of LC-MS data in cooperation with XCMS (158, 159). Online MS/MS fragmentation, UV spectra, and estimates of partition coefficients based on retention time have been used to further investigate metabolite structures (157, 160, 161). The MagMA software package recently launched is able to read multistage tandem mass spectral data to add potential candidates based on in silico–predicted fragmentation (162). In particular, the use of accurate fragmentation mass data as input can enhance the metabolite identification process by selecting the most likely candidates on the basis of similarities in fragmentation pathways and their readily assigned elemental formula with the unknown query metabolite as exemplified by

dietary polyphenols (144). A large number of polyphenol metabolites such as glucuronides of 5-(3′,4′-dihydroxyphenyl)-γ-valerolactone and sulfate esters of methylated (epi)catechin could thus be easily annotated and some fully structurally elucidated by using a combination of MS fragmentation and NMR (163).

Moreover, recently developed bioinformatics approaches aim to narrow the number of possible candidate structures that match with an unknown query metabolite by taking into account the chemical and biological background of the sample (164). For example, it is more likely that a metabolite excreted in urine is more polar as a result of phase II reactions. This has predictable consequences for its expected mass and chromatographic behavior, which can be used to mine metabolomic data sets. It is expected that these various software tools will be beneficial in the hunt for metabolite entities represented by the food metabolome.

PERSPECTIVES FOR FUTURE APPLICATIONS OF THE FOOD METABOLOME

Discovering disease-related dietary factors

MWASs have been proposed as useful tools for discovering low-molecular-weight biomarkers that are predictive of either causal exposures or disease progression (59, 165, 166). In fact, MWASs can be regarded as a special case of the exposome-wide association study, which investigates disease associations with all exposures to low- and high-molecular-weight compounds (167). Given the thousands of potentially important exposures to consider, MWASs and exposome-wide association studies move away from knowledge-driven designs that focus on a priori hypotheses about particular exposures toward data-driven designs using untargeted or semitargeted sets of analytes (167). In either case, potentially useful biomarkers may be identified through rigorous comparisons of quantitative or semiquantitative profiles of biospecimens obtained from subjects with and without a particular disease (59). Because diets and lifestyle strongly affect the metabolome, any pending disease may lead to reverse causation in MWASs; study design and interpretation must therefore take into account the common responses to early signs of disease in the population under study and other potential confounders.

This biomarker discovery process is shown in **Figure 2**. With a focus on the food metabolome and associated biomarkers of potentially causal dietary exposures, the figure includes both semitargeted and untargeted designs. In the semitargeted approach, preliminary cross-sectional studies are developed to connect dietary records with the food metabolome and thereby identify dietary biomarkers that are highly correlated with the consumption of particular foods. A good example of this approach is given by Saadatian-Elahi et al (168), who correlated food consumption, as determined by 24-h dietary recall, with plasma concentrations of 22 fatty acids determined by gas chromatography in 3000 subjects from the European Prospective Investigation into Cancer and Nutrition cohort. Strong correlations between regional dietary factors and fatty acid concentrations allowed components of the food metabolome to be used as predictor variables in a prospective investigation of gastric cancer in the European Prospective Investigation into Cancer and Nutrition cohort. Three fatty acids—oleic acid, α-linolenic acid, and di-homo-γ-linolenic acid—were found to be associated with the risk of gastric cancer (169). These associations were tentatively explained by either different amounts of dietary intake or differential fatty acid metabolism in cases and controls.

The alternative untargeted approach makes no a priori assumptions regarding sources of exposure that are causal for a particular disease but instead relies on comparisons of comprehensive profiles of metabolomic features between cases and controls to find discriminating exposure biomarkers. Once these exposure biomarkers have been identified, follow-up studies are performed to determine their sources (167), and those related to dietary factors would be regarded as disease-associated dietary biomarkers (Figure 2). The agnostic nature of the untargeted design allows all potentially useful biomarkers to be identified, including not only dietary biomarkers but also those related to endogenous factors (including the microbiota), pollution, and drugs as well as biomarkers of disease progression. A good example of the untargeted approach is given by Holmes et al (59) and Bictash et al (166) who used untargeted NMR of >4000 urine specimens from the INTERMAP study to investigate potentially causal factors for high blood pressure across geographically diverse populations. The investigators showed that metabolite concentrations differed substantially between Asian and Western populations, suggesting important effects of diet and related risk factors, including the microbiota, on the risk of coronary artery disease and stroke. Three highly discriminating biomarkers were identified, namely alanine, which was directly correlated with blood pressure, and formate and hippurate, both of which were inversely correlated with blood pressure. All of these discriminatory biomarkers point to dietary sources, sometimes in combination with cometabolism by gut microbiota. For example, alanine is associated with diets that emphasize animal products rather than vegetables, and hippurate has been associated with microbiota colonization of the gut (170).

A more recent example of the untargeted approach is provided by a series of articles from Stanley Hazen's group at the Cleveland Clinic (108, 171, 172). In their initial untargeted LC-MS/MS investigation (171), the authors showed that the nutrient choline, along with its major metabolites, betaine and TMAO, were associated with risks of cardiovascular disease, particularly TMAO. Then, by using an elegant set of targeted follow-up

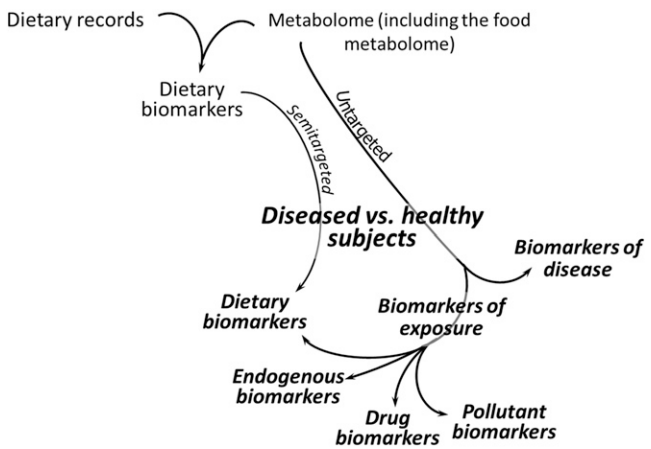

FIGURE 2. The food metabolome and discovery of food-related biomarkers associated with diseases. Both semitargeted and untargeted approaches are shown. Disease-validated biomarkers are shown in bold letters.

studies with the use of a choline challenge as well as characterization and manipulation of microbiota, Hazen and coworkers showed that consumption of foods rich in choline or carnitine, such as eggs, milk, liver, meat, or fish, produced high concentrations of TMAO in both humans and animals who possessed the requisite microbiota for metabolizing choline or carnitine to trimethylamine, the immediate precursor of TMAO (108, 171, 172). Subjects from a cohort of >2000 patients with cardiovascular disease, who were in the highest quartile for plasma TMAO concentrations, had a >2-fold risk of a heart attack or stroke compared with subjects from the lowest quartile (108, 172).

Identification of new potentially bioactive food–derived metabolites

The application of metabolomics to foods has allowed the identification of a large variety of novel food constituents that are either naturally present in the food species or formed during food processing (65, 173, 174). Similarly, the exploration of the food metabolome in human biofluids by means of wide-coverage profiling methods and intermetabolite correlation analysis (175) should show exposures to many nonnutrient food compounds and their metabolites whose presence has not been previously identified. These compounds could also be new bioactive compounds. As an example, the recent description of benzoxazinoids in rye facilitated the identification of some of their metabolites (2,4-dihydroxy-1,4-benzoxazin-3-one, 2-aminophenol sulfate, and hydroxylated phenylacetamides) in urinary metabolic profiles observed after rye bread consumption (86, 114). These benzoxazinoid metabolites certainly deserve further investigation as potential contributors to the health effects of rye products because of some documented anti-inflammatory, immunoregulatory, and appetite-suppressing properties (176). This example shows that a better characterization of chemicals contained in a given food should markedly improve our understanding of food-derived exposures and their biological effects.

Metabolomics will help nutrition researchers move away from the reductionist views on health effects of foods that have largely prevailed until today. For many years, health effects associated with a particular food have often been attributed to just 1 or 2 of their constituents on the basis of certain biological properties observed in vitro. Examples include lycopene in tomato, which is thought to prevent prostate cancer; isoflavones in soy products, which may prevent hormone-dependent cancers; and catechins in tea or flavanones in citrus, which may play a role in the prevention of cardiovascular diseases. Although these compounds may actually contribute to the health effects of the food, as has been well demonstrated in intervention studies in which a whole food has been compared with one of its bioactive constituents (177), their popularity may have overshadowed the contribution of other, lesser known constituents also present in the same food. Metabolomics could potentially reveal these other bioactive constituents, and the approach is already being used in the characterization of multicomponent drugs and herbal medicines (178). Knowledge of all circulating metabolites is essential to understand the effects of the diet on health, and new metabolites formed from nutrients and other food constituents are continuously being identified, even for widely studied compounds (101, 179).

The combining of metabolomic with genomic data will also be important to identify dietary compounds causally related to diseases. A number of loci associated with variations in the concentration of endogenous metabolites could be identified in genome-wide analysis studies, and many of these genes were coding for metabolic enzymes (180–182). Various enzymes involved in the biotransformation of xenobiotics and dietary compounds also show genetic polymorphisms (183, 184). The analysis of their variants combined with that of the food metabolome in MWASs may reinforce the associations observed between food-derived metabolites and disease risk as has been previously found for alcohol or folate as disease risk factors (185, 186). A deeper knowledge of the enzymes involved in the biotransformation of dietary compounds is, however, needed to warrant success of this approach.

Another possible approach to identify food compounds potentially responsible for the activity is the study of longitudinal variations in their concentrations and their associations with particular health outcomes or surrogate health markers in population studies or clinical trials. The kinetics of a metabolite's appearance in plasma after a meal can be related to the kinetics of associated physiologic events. Epicatechin metabolite concentrations in plasma after cocoa intake paralleled the increase of plasma nitroso species concentrations and the vascular response (187). Overall, the study of the variability in the food metabolome (which permeates all human tissues) and its association with health outcomes should greatly contribute to the identification of the food metabolites responsible for the effects of diet on health and diseases.

NETWORKING AND RECOMMENDATIONS TO MOVE THE FIELD FORWARD

As detailed in the previous sections, food metabolomics requires inputs from specialists from various disciplines, including analytic chemistry, chemometrics, statistics, bioinformatics, nutritional science, and biology. Within one group, it is difficult to cover all the techniques and methods required to perform a comprehensive metabolomic study. Several networking initiatives may help in this respect by providing rapid access to new information and tools. The rapid pace of development in metabolomic profiling techniques makes the role of networks even more important to help absorb and facilitate the use of all of the information. This is supported by creating databases for compound information and spectral data, libraries of chemical standards, algorithms for data analysis, repositories for raw data and metadata, and standardization initiatives to define current good practices. The Metabolomics Standards Initiative (MSI) launched by the Metabolomics Society is an example of such an initiative, and the MSI has already had significant impact on reporting formats in metabolomics (110, 188).

Further networking initiatives to share knowledge in open discussion or workshops, the sharing of data and standard operating protocols, as well as starting common training initiatives will be important to accelerate progress in the field. A good example of such a sharing network is the European Nutrigenomics Organization, a not-for-profit private organization with academic and private institutional members from all over the world. None of these efforts should be seen as static but rather as a current collective instrument to help in the release of biological information at the level of the metabolome. There are several open-user forums working with the development and application of metabolomic software and

standards in general, such as the MSI working groups (189), the Metabolomics Forum (190), and several others, but none of these relate specifically to the food metabolome. The First International Workshop on the Food Metabolome was a first occasion for all researchers active in the field to meet and make propositions for future research. These propositions are summarized here.

Coordination of dietary studies

The food metabolome is exceedingly complex because it encompasses metabolites derived from as many metabolomes as there are edible species. Therefore, a particularly focused community effort is necessary to reach our ultimate goal of full coverage for all foods and all food metabolites. A large number of studies with different designs will be necessary to validate each dietary marker. For example, many studies have been conducted with oranges (Table 3), but a broad coverage of all citrus and many other fruit as well as kinetic studies have been necessary to interpret proline betaine as a short-term marker of citrus that is dominated by orange and orange juice intake (80). Similar work is needed and could be a shared effort for many other food groups, including cruciferous and apiaceous vegetables, pomes, cheeses, meats, fish, and others.

A large concerted action or open-project network would be needed to help prioritize needs for novel markers and focus on areas in which drugs have largely failed and where diet and nutrition show promise to prevent or cure diseases. More discussion is clearly needed between laboratory scientists, nutritionists, and epidemiologists to address this question in a rational way. Such a network might share information on current research plans to avoid redundancy, share known as well as unidentified markers related to specific foods, or even form a shared workflow pipeline for dietary studies, data analysis, and metabolite identification. In addition, the constitution of a database describing resources of high-quality human samples collected in various dietary intervention studies developed for other purposes would also be extremely useful. This information is partly accessible in a database such as ClinicalTrials.gov (191), but no indication is given on the availability of biospecimens. These samples would prove very useful for biomarker validation purposes and would save a lot of effort and money otherwise needed to replicate such clinical studies. An example of a local, but open, sample repository for experimental studies including nutrition is the CUBE biobank, which covers samples from a single university (www.cube.ku.dk). An umbrella of such local repositories could be one possible way forward to improve reuse of samples for biomarker validation studies.

Software tools

A comprehensive set of software tools has been developed and shared to help the scientific community that covers every step in data processing and analysis. Most of these are not specific to the food metabolome analysis (*see* Analysis of metabolomic data and Software tools for annotation of the food metabolome sections). However, for the identification of food-derived metabolites, additional software developments are needed, particularly for in silico prediction of the metabolism of compounds found in foods. Some commercial software exists for the pharmaceutical area (192, 193) and covers many phase I and II reactions. However, many compounds in foods have structures that are uncommon in pharmaceuticals. Food constituents may be degraded by specialized enzymes and may also be extensively metabolized by the gut microbiota, so these metabolic pathways need to be covered as well. This work will require a large community effort to develop software to predict structures from all possible metabolites from any food compound, including their conjugates; such a tool could be further combined with software that performs in silico fragmentation to predict daughter ions and additional prediction tools for predicting physicochemical properties such as polarity and hence retention time. Prediction of absorption, distribution, and excretion of the food compounds and of their metabolites would be an additional area that would help the food metabolome community. Systematic in silico–predicted metabolites could also be stored in food metabolome databases.

Databases

The human metabolome database has recently expanded to include compounds found in common foods because these are, at least initially before metabolism, also present within the human body (9). Databases specific for the food metabolome are still largely missing apart from Phenol-Explorer, a database on all known polyphenol metabolites (70). The development of similar databases for other classes of food compounds will likely require a coordinated effort from many researchers active in various fields. These databases should provide spectral data for the food-derived metabolites in each class and any information useful for their identification. When not available, in silico–predicted mass fragmentation spectra could be calculated and also stored, as is done in SciFinder (151). The same databases could additionally allow metabolites to be linked with their food precursors, as well as with their possible dietary sources (70). The involvement of food scientists will be essential to provide this information.

Study repositories with processed metabolomic data

To shape consensus and create openness in the evolving field of metabolomics, it is important to share data and information on food metabolome studies, as is done in many other biomedical fields (194–196). Indeed, for many funding agencies, this is becoming a key condition of funding. One such initiative is the Metabolights database, which aims to shape a fully open-access, shared database for metabolomic studies (197). Raw and processed data and metadata can be uploaded and curated before deposition into the Metabolights core database, which then makes the information accessible through the Internet. A similar ongoing but conceptually broader initiative is the Nutritional Phenotype Database (dbNP) (198), initiated by the Nutrigenomics Organization. The dbNP can hold data from several omics platforms, including metabolomics, together with study metadata in a searchable format. It is open access and builds on private accounts for uploading and analyzing data with the possibility of open sharing when data can be released for others. Both dbNP and Metabolights provide several online software tools to help in data curation and analysis.

The storage of searchable, annotated, raw analytic data files with well-documented dietary metadata from human intervention or cross-sectional studies will facilitate the comparison of raw or preprocessed data with previously obtained spectral data of food-derived metabolites. Such a repository that contains all unknowns detected in previous food metabolome studies would be a precious aid to identify the most robust dietary biomarkers. The

format of raw analytic data concerning the food metabolome as well as that of the dietary metadata will have to be defined.

Food metabolome reference library

The definitive identification of biomarkers is often hampered by the lack of available chemical standards. The large majority of the components of the food metabolome are not commercially available. The development of a resource to synthesize and distribute chemical standards should be a priority. The development of a shared or federated resource of chemical standards for dietary metabolites will allow researchers to confirm or validate compound identifications. Food scientists and natural product chemists who have isolated from various foods and related products or synthesized these chemicals should be associated with this effort. Biotransformation routes (enzymes, microorganisms) could also be better exploited particularly to synthesize conjugated metabolites.

Standardization initiatives

The MSI, initiated by the Metabolomics Society, has already issued several reference articles on good practice for metabolomic research. The MSI is broad and includes activities by several working groups covering many aspects of metabolomics (189). However, there is no current standardization initiative for the food metabolome, and the current article is launched as a starting signal for such an initiative to share tools, information, and data sets to help encourage a community-driven advancement of research on the food metabolome.

One of the most needed initiatives is a strategy with associated tools for the validation of food intake biomarkers. In particular, a reference database could integrate all potential markers that can be included into targeted multimetabolite methods; each marker listed could be graded indicating the level of validation: for example, "0" for a newly discovered putative biomarker, "1" for a biomarker with analytic validation including kinetics in the sample type of interest, "2" for a biomarker also validated in a controlled dietary intervention studies as well as in cross-sectional studies, and "3" for a biomarker also confirmed to be in accordance with other markers for the same food or foods.

Training

Finally, education and training in the food metabolome also represent an effort that needs to be shared at all levels, from creating the basic training texts to specialized courses at several academic levels. Several training workshops on metabolomics and nutrition were supported by the Nutrigenomics Organization (112). However, more sustained and focused support is needed to bring the field of food metabolome research to maturity.

CONCLUSIONS

Recent analyses of the food metabolome with modern analytic and bioinformatic tools have shown the considerable extent of information on dietary exposure contained in human biospecimens. Some proof-of-principle studies have established the feasibility of metabolomic approaches to identify novel dietary biomarkers and suggest that these approaches could greatly expand the field of molecular nutritional epidemiology and contribute to future progress in nutrition research. Propositions made here to define common objectives and priorities, optimize study designs, develop databases and software tools, and promote sharing of data and resources should contribute to bringing this emerging field to maturity. A dialogue between nutritionists, epidemiologists, analysts, chemometricians, statisticians, and bioinformaticians has just begun. It will be essential to build multidisciplinary projects and make sure that the design of future studies is defined and optimized to answer to nutritionists' and epidemiologists' most urgent needs for biomarkers. Major progress in assessing complex dietary exposures at the individual level is expected from these biomarkers. They should also significantly contribute to a better understanding of the complex interactions between diet and human health.

We thank Mazda Jenab (International Agency for Research on Cancer, Lyon) for helpful comments on the manuscript. We also thank all of the participants in the First International Workshop on the Food Metabolome for their active participation in the discussions from which many of the recommendations presented in this article were derived.

The authors' responsibilities were as follows—AS, CM, JD, and LB: contributed to the conception and design of the manuscript; and AS provided extensive feedback concerning all versions of the manuscript and had primary responsibility for final content. All of the authors contributed to the writing of separate sections of the manuscript and reviewed and edited all versions of the manuscript. All of the authors read and approved the final manuscript. None of the authors had a conflict of interest.

REFERENCES

1. Doets EL, de Wit LS, Dhonukshe-Rutten RAM, Cavelaars A, Raats MM, Timotijevic L, Brzozowska A, Wijnhoven TMA, Pavlovic M, Totland TH, et al. Current micronutrient recommendations in Europe: towards understanding their differences and similarities. Eur J Nutr 2008;47:17–40.
2. Dunn WB, Broadhurst DI, Atherton HJ, Goodacre R, Griffin JL. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. Chem Soc Rev 2011;40:387–426.
3. Wild CP, Scalbert A, Herceg Z. Measuring the exposome: a powerful basis for evaluating environmental exposures and cancer risk. Environ Mol Mutagen 2013;54:480–99.
4. Rezzi S, Ramadan Z, Fay LB, Kochhar S. Nutritional metabonomics: applications and perspectives. J Proteome Res 2007;6:513–25.
5. Wishart DS. Metabolomics: applications to food science and nutrition research. Trends Food Sci Technol 2008;19:482–93.
6. Manach C, Hubert J, Llorach R, Scalbert A. The complex links between dietary phytochemicals and human health deciphered by metabolomics. Mol Nutr Food Res 2009;53:1303–15.
7. Fardet A, Llorach R, Orsoni A, Martin JF, Pujos-Guyot E, Lapierre C, Scalbert A. Metabolomics provide new insights on the metabolism of dietary phytochemicals in rats. J Nutr 2008;138:1282–7.
8. Cevallos-Cevallos JM, Reyes-De-Corcuera JI, Etxeberria E, Danyluk MD, Rodrick GE. Metabolomic analysis in food science: a review. Trends Food Sci Technol 2009;20:557–66.
9. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, et al. HMDB 3.0—the Human Metabolome Database in 2013. Nucleic Acids Res 2013;41 (Database issue):D801–7.
10. Rutishauser IHE. Dietary intake measurements. Public Health Nutr 2005;8:1100–7.
11. Bingham S, Luben R, Welch A, Tasevska N, Wareham N, Khaw KT. Epidemiologic assessment of sugars consumption using biomarkers: comparisons of obese and nonobese individuals in the European Prospective Investigation of Cancer Norfolk. Cancer Epidemiol Biomarkers Prev 2007;16:1651–4.
12. Marshall JR, Chen Z. Diet and health risk: risk patterns and disease-specific associations. Am J Clin Nutr 1999;69:1351S–6S.

13. Jenab M, Slimani N, Bictash M, Ferrari P, Bingham SA. Biomarkers in nutritional epidemiology: applications, needs and new horizons. Hum Genet 2009;125:507–25.
14. Potischman N, Freudenheim JL. Biomarkers of nutritional exposure and nutritional status: an overview. J Nutr 2003;133(suppl):873S–4S.
15. London SJ, Yuan JM, Chung FL, Gao YT, Coetzee GA, Ross RK, Yu MC. Isothiocyanates, glutathione S-transferase M1 and T1 polymorphisms, and lung-cancer risk: a prospective study of men in Shanghai, China. Lancet 2000;356:724–9.
16. Kensler TW, Roebuck BD, Wogan GN, Groopman JD. Aflatoxin: a 50-year odyssey of mechanistic and translational toxicology. Toxicol Sci 2011;120:S28–48.
17. Baldrick FR, Woodside JV, Elborn JS, Young IS, McKinley MC. Biomarkers of fruit and vegetable intake in human intervention studies: a systematic review. Crit Rev Food Sci Nutr 2011;51:795–815.
18. Mennen LI, Sapinho D, Ito H, Galan P, Hercberg S, Scalbert A. Urinary flavonoids and phenolic acids as biomarkers of intake for polyphenol-rich foods. Br J Nutr 2006;96:191–8.
19. Ross AB, Bourgeois A, Macharia HN, Kochhar S, Jebb SA, Brownlee IA, Seal CJ. Plasma alkylresorcinols as a biomarker of whole-grain food consumption in a large population: results from the WHOLEheart Intervention Study. Am J Clin Nutr 2012;95:204–11.
20. Andersson A, Marklund M, Diana M, Landberg R. Plasma alkylresorcinol concentrations correlate with whole grain wheat and rye intake and show moderate reproducibility over a 2-to 3-month period in free-living Swedish adults. J Nutr 2011;141:1712–8.
21. Verkasalo PK, Appleby PN, Allen NE, Davey G, Adlercreutz H, Key TJ. Soya intake and plasma concentrations of daidzein and genistein: validity of dietary assessment among eighty British women (Oxford arm of the European Prospective Investigation into Cancer and Nutrition). Br J Nutr 2001;86:415–21.
22. Allen NE, Grace PB, Ginn A, Travis RC, Roddam AW, Appleby PN, Key T. 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;99:653–9.
23. Myint T, Fraser GE, Lindsted KD, Knutsen SF, Hubbard RW, Bennett HW. Urinary 1-methylhistidine is a marker of meat consumption in black and in white California Seventh-day Adventists. Am J Epidemiol 2000;152:752–5.
24. Arsenault LN, Matthan N, Scott TM, Dallal G, Lichtenstein AH, Folstein MF, Rosenberg I, Tucker KL. Validity of estimated dietary eicosapentaenoic acid and docosahexaenoic acid intakes determined by interviewer-administered food frequency questionnaire among older adults with mild-to-moderate cognitive impairment or dementia. Am J Epidemiol 2009;170:95–103.
25. Hodgson JM, Chan SY, Puddey IB, Devine A, Wattanapenpaiboon N, Wahlqvist ML, Lukito W, Burke V, Ward NC, Prince RL, et al. Phenolic acid metabolites as biomarkers for tea- and coffee-derived polyphenol exposure in human subjects. Br J Nutr 2004;91:301–6.
26. Brantsaeter AL, Haugen M, Julshamn K, Alexander J, Meltzer HM. Evaluation of urinary iodine excretion as a biomarker for intake of milk and dairy products in pregnant women in the Norwegian Mother and Child Cohort Study (MoBa). Eur J Clin Nutr 2009;63:347–54.
27. Turunen AW, Mannisto S, Kiviranta H, Marniemi J, Jula A, Tiittanen P, Suominen-Taipale L, Vartiainen T, Verkasalo PK. Dioxins, polychlorinated biphenyls, methyl mercury and omega-3 polyunsaturated fatty acids as biomarkers of fish consumption. Eur J Clin Nutr 2010;64:313–23.
28. Kroke A, Klipstein-Grobusch K, Hoffmann K, Terbeck I, Boeing H, Helander A. Comparison of self-reported alcohol intake with the urinary excretion of 5-hydroxytryptophol: 5-hydroxyindole-3-acetic acid, a biomarker of recent alcohol intake. Br J Nutr 2001;85:621–7.
29. Crews H, Alink G, Andersen R, Braesco V, Holst B, Maiani G, Ovesen L, Scotter M, Solfrizzo M, van den Berg R, et al. A critical assessment of some biomarker approaches linked with dietary intake. Br J Nutr 2001;86(suppl 1):S5–35.
30. Vioque J, Weinbrenner T, Asensio L, Castello A, Young IS, Fletcher A. Plasma concentrations of carotenoids and vitamin C are better correlated with dietary intake in normal weight than overweight and obese elderly subjects. Br J Nutr 2007;97:977–86.
31. Brown MJ, Ferruzzi MG, Nguyen ML, Cooper DA, Eldridge AL, Schwartz SJ, White WS. Carotenoid bioavailability is higher from salads ingested with full-fat than with fat-reduced salad dressings as measured with electrochemical detection. Am J Clin Nutr 2004;80:396–403.
32. Albanes D, Virtamo J, Taylor PR, Rautalahti M, Pietinen P, Heinonen OP. Effects of supplemental beta-carotene, cigarette smoking, and alcohol consumption on serum carotenoids in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study. Am J Clin Nutr 1997;66:366–72.
33. Zamora-Ros R, Touillaud M, Rothwell JA, Romieu I, Scalbert A. Measuring exposure to the polyphenol metabolome in epidemiological studies: current tools, applications and their limits. Am J Clin Nutr (in press).
34. Manach C, Williamson G, Morand C, Scalbert A, Remesy C. Bioavailability and bioefficacy of polyphenols in humans. I. Review of 97 bioavailability studies. Am J Clin Nutr 2005;81:230S–42S.
35. Söderholm PP, Koskela AH, Lundin JE, Tikkanen MJ, Adlercreutz HC. Plasma pharmacokinetics of alkylresorcinol metabolites: new candidate biomarkers for whole-grain rye and wheat intake. Am J Clin Nutr 2009;90:1167–71.
36. Burri BJ, Neidlinger TR, Clifford AJ. Serum carotenoid depletion follows first-order kinetics in healthy adult women fed naturally low carotenoid diets. J Nutr 2001;131:2096–100.
37. Kumar A, Vineis P, Sacerdote C, Fiorini L, Sabbioni G. Determination of new biomarkers to monitor the dietary consumption of isothiocyanates. Biomarkers 2010;15:739–45.
38. Outzen M, Egeberg R, Dragsted L, Christensen J, Olesen PT, Frandsen H, Overvad K, Tjonneland A, Olsen A. Dietary determinants for Hb-acrylamide and Hb-glycidamide adducts in Danish non-smoking women. Br J Nutr 2011;105:1381–7.
39. Grigoryan H, Li H, Iavarone AT, Williams ER, Rappaport SM. Cys34 adducts of reactive oxygen species in human serum albumin. Chem Res Toxicol 2012;25:1633–42.
40. Rappaport SM, Li H, Grigoryan H, Funk WE, Williams ER. Adductomics: characterizing exposures to reactive electrophiles. Toxicol Lett 2012;213:83–90.
41. Pérez-Jiménez J, Hubert J, Ashton K, Hooper L, Cassidy A, Manach C, Williamson G, Scalbert A. Urinary metabolites as biomarkers of polyphenol intake in humans—a systematic review. Am J Clin Nutr 2010;92:801–9.
42. Levine M, Conry-Cantilena C, Wang Y, Welch RW, Washko PW, Dhariwal KR, Park JB, Lazarev A, Graumlich JF, King J, et al. Vitamin C pharmacokinetics in healthy volunteers: evidence for a recommended dietary allowance. Proc Natl Acad Sci USA 1996;93:3704–9.
43. Pirro V, Valente V, Oliveri P, De Bernardis A, Salomone A, Vincenti M. Chemometric evaluation of nine alcohol biomarkers in a large population of clinically-classified subjects: pre-eminence of ethyl glucuronide concentration in hair for confirmatory classification. Anal Bioanal Chem 2011;401:2153–64.
44. Lloyd AJ, Beckmann M, Fave G, Mathers JC, Draper J. Proline betaine and its biotransformation products in fasting urine samples are potential biomarkers of habitual citrus fruit consumption. Br J Nutr 2011;106:812–24.
45. Al-Delaimy WK, Ferrari P, Slimani N, Pala V, Johansson I, Nilsson S, Mattisson I, Wirfalt E, Galasso R, Palli D, et al. Plasma carotenoids as biomarkers of intake of fruits and vegetables: individual-level correlations in the European Prospective Investigation into Cancer and Nutrition (EPIC). Eur J Clin Nutr 2005;59:1387–96.
46. Scalbert A, Manach C, Morand C, Rémésy C, Jiménez L. Dietary polyphenols and the prevention of diseases. Crit Rev Food Sci Nutr 2005;45:287–306.
47. Medina-Remón A, Zamora-Ros R, Rotchés-Ribalta M, Andres-Lacueva C, Martínez-González MA, Covas MI, Corella D, Salas-Salvadó J, Gómez-Gracia E, Ruiz-Gutiérrez V, et al. Total polyphenol excretion and blood pressure in subjects at high cardiovascular risk. Nutr Metab Cardiovasc Dis 2011;21:323–31.
48. Ross AB, Shepherd MJ, Schupphaus M, Sinclair V, Alfaro B, Kamal-Eldin A, Aman P. Alkylresorcinols in cereals and cereal products. J Agric Food Chem 2003;51:4111–8.
49. Pujos-Guillot E, Hubert J, Martin J-F, Lyan B, Quintana M, Claude S, Chabanas B, Rothwell JA, Bennetau-Pelissero C, Scalbert A, 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;12:1645–59.

The American Journal of Clinical Nutrition

50. van der Werf MJ, Overkamp KM, Muilwijk B, Coulier L, Hankemeier T. Microbial metabolomics: toward a platform with full metabolome coverage. Anal Biochem 2007;370:17–25.
51. Bais P, Moon SM, He K, Leitao R, Dreher K, Walk T, Sucaet Y, Barkan L, Wohlgemuth G, Roth MR, et al. PlantMetabolomics.org: a web portal for plant metabolomics experiments. Plant Physiol 2010; 152:1807–16.
52. Wishart DS. Human Metabolome Database: completing the 'human parts list'. Pharmacogenomics 2007;8:683–6.
53. Takeda I, Stretch C, Barnaby P, Bhatnager K, Rankin K, Fu H, Weljie A, Jha N, Slupsky C. Understanding the human salivary metabolome. NMR Biomed 2009;22:577–84.
54. Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, Sinelnikov I, Krishnamurthy R, Eisner R, Gautam B, et al. The human serum metabolome. PLoS ONE 2011;6:e16957.
55. Solanky KS, Bailey NJ, Beckwith-Hall BM, Bingham S, Davis A, Holmes E, Nicholson JK, Cassidy A. Biofluid 1H NMR-based metabonomic techniques in nutrition research—metabolic effects of dietary isoflavones in humans. J Nutr Biochem 2005;16:236–44.
56. Tulipani S, Llorach R, Jáuregui O, López-Uriarte P, Garcia-Aloy M, Bullo M, Salas-Salvadó J, Andrés-Lacueva C. Metabolomics unveils urinary changes in subjects with metabolic syndrome following 12-week nut consumption. J Proteome Res 2011;10:5047–58.
57. Cross AJ, Major JM, Sinha R. Urinary biomarkers of meat consumption. Cancer Epidemiol Biomarkers Prev 2011;20:1107–11.
58. Van Dorsten FA, Daykin CA, Mulder TPJ, Van Duynhoven JPM. Metabonomics approach to determine metabolic differences between green tea and black tea consumption. J Agric Food Chem 2006;54: 6929–38.
59. Holmes E, Loo RL, Stamler J, Bictash M, Yap IKS, Chan Q, Ebbels T, De Iorio M, Brown IJ, Veselkov KA, et al. Human metabolic phenotype diversity and its association with diet and blood pressure. Nature 2008;453:396–400.
60. Slupsky CM, Rankin KN, Wagner J, Fu H, Chang D, Weljie AM, Saude EJ, Lix B, Adamko DJ, Shah S, et al. Investigations of the effects of gender, diurnal variation, and age in human urinary metabolomic profiles. Anal Chem 2007;79:6995–7004.
61. Krug S, Kastenmueller G, Stueckler F, Rist MJ, Skurk T, Sailer M, Raffler J, Roemisch-Margl W, Adamski J, Prehn C, et al. The dynamic range of the human metabolome revealed by challenges. FASEB J 2012;26:2607–19.
62. Heinzmann SS, Merrifield CA, Rezzi S, Kochhar S, Lindon JC, Holmes E, Nicholson JK. Stability and robustness of human metabolic phenotypes in response to sequential food challenges. J Proteome Res 2012;11:643–55.
63. Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, Custodio DE, Abagyan R, Siuzdak G. METLIN—a metabolite mass spectral database. Ther Drug Monit 2005;27:747–51.
64. Fiehn O. Metabolomics—the link between genotypes and phenotypes. Plant Mol Biol 2002;48:155–71.
65. Iijima Y, Nakamura Y, Ogata Y, Tanaka K, Sakurai N, Suda K, Suzuki T, Suzuki H, Okazaki K, Kitayama M, et al. Metabolite annotations based on the integration of mass spectral information. Plant J 2008;54:949–62.
66. Guo AC, Jewison T, Wilson M, Liu Y, Knox C, Djoumbou Y, Lo P, Mandal R, Krishnamurthy R, Wishart DS. ECMDB: the E-coli Metabolome Database. Nucleic Acids Res 2013;41:D625–30.
67. Nicholson JK, Holmes E, Kinross J, Burcelin R, Gibson G, Jia W, Pettersson S. Host-gut microbiota metabolic interactions. Science 2012;336:1262–7.
68. Testa B, Pedretti A, Vistoli G. Foundation review: reactions and enzymes in the metabolism of drugs and other xenobiotics. Drug Discov Today 2012;17:549–60.
69. Langowski J, Long A. Computer systems for the prediction of xenobiotic metabolism. Adv Drug Deliv Rev 2002;54:407–15.
70. Rothwell JR, Urpi-Sarda M, Boto-Ordonez M, Knox C, Llorach R, Eisner R, Cruz J, Neveu V, Wishart D, Manach C, et al. Phenol-Explorer 2.0: a major update of the Phenol-Explorer database integrating data on polyphenol metabolism and pharmacokinetics in humans and experimental animals. Database 2012;2012:bas031.
71. Silberberg M, Morand C, Mathevon T, Besson C, Manach C, Scalbert A, Remesy C. The bioavailability of polyphenols is highly governed by the capacity of the intestine and of the liver to secrete conjugated metabolites. Eur J Nutr 2006;45:88–96.
72. Peris-Bondia F, Latorre A, Artacho A, Moya A, D'Auria G. The active human gut microbiota differs from the total microbiota. PLoS ONE 2011;6:e22448.
73. Moco S, Martin F-P, Rezzi S. A metabolomics view on gut microbiome modulation by polyphenol-rich foods. J Proteome Res 2012; 11:4781–90.
74. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto J-M, et al. Enterotypes of the human gut microbiome. Nature 2011;473:174–80.
75. Jewison T, Knox C, Neveu V, Djoumbou Y, Guo AC, Lee J, Liu P, Mandal R, Krishnamurthy R, Sinelnikov I, et al. YMDB: the Yeast Metabolome Database. Nucleic Acids Res 2012;40:D815–20.
76. Wishart DS. Systems biology resource arising from the Human Metabolome Project. In: Suhre K, ed. Genetics meets metabolomics. New York, NY: Springer, 2012:157–75.
77. Institut National de la Recherche Agronomique. PhytoHub. Available from: www.phytohub.eu (cited 3 September 2013).
78. University of Alberta. FooDB: the food component database. Available from: http://www.foodb.ca/ (cited 3 September 2013).
79. Chajès V, Biessy C, Byrnes G, Deharveng G, Saadatian-Elahi M, Jenab M, Peeters PHM, Ocke M, Bueno-de-Mesquita HB, Johansson I, et al. Ecological-level associations between highly processed food intakes and plasma phospholipid elaidic acid concentrations: results from a cross-sectional study within the European Prospective Investigation into Cancer and Nutrition (EPIC). Nutr Cancer 2011;63: 1235–50.
80. Heinzmann SS, Brown IJ, Chan Q, Bictash M, Dumas M-E, Kochhar S, Stamler J, Holmes E, Elliott P, Nicholson JK. Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. Am J Clin Nutr 2010;92:436–43.
81. Andersen MB, Kristensen M, Manach C, Pujos-Guillot E, Poulsen SK, Larsen TM, Astrup A, Dragsted L. Discovery and validation of urinary exposure markers for different plant foods by untargeted metabolomics. Anal Bioanal Chem 2014;406:1829–44.
82. Lloyd AJ, Favé G, Beckmann M, Lin W, Tailliart K, Xie L, Mathers JC, Draper J. Use of mass spectrometry fingerprinting to identify urinary metabolites after consumption of specific foods. Am J Clin Nutr 2011;94:981–91.
83. O'Sullivan A, Gibney MJ, Brennan L. Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. Am J Clin Nutr 2011;93:314–21.
84. Edmands WMB, Beckonert OP, Stella C, Campbell A, Lake BG, Lindon JC, Holmes E, Gooderham NJ. Identification of human urinary biomarkers of cruciferous vegetable consumption by metabonomic profiling. J Proteome Res 2011;10:4513–21.
85. Andersen M-B, Reinbach H, Rinnan A, Barri T, Mithril C, Dragsted L. Discovery of exposure markers in urine for Brassica-containing meals served with different protein sources by UPLC-qTOF-MS untargeted metabolomics. Metabolomics 2013;9:984–97.
86. Bondia-Pons I, Barri T, Hanhineva K, Juntunen K, Dragsted LO, Mykkänen H, Poutanen K. UPLC-QTOF/MS metabolic profiling unveils urinary changes in humans after a whole grain rye versus refined wheat bread intervention. Mol Nutr Food Res 2013;57:412–22.
87. Lloyd AJ, Beckmann M, Haldar S, Seal C, Brandt K, Draper J. Data-driven strategy for the discovery of potential urinary biomarkers of habitual dietary exposure. Am J Clin Nutr 2013;97:377–89.
88. Lang R, Wahl A, Stark T, Hofmann T. Urinary N-methylpyridinium and trigonelline as candidate dietary biomarkers of coffee consumption. Mol Nutr Food Res 2011;55:1613–23.
89. Wang Y, Tang H, Nicholson JK, Hylands PJ, Sampson J, Holmes E. A metabonomic strategy for the detection of the metabolic effects of chamomile (Matricaria recutita L.) ingestion. J Agric Food Chem 2005;53:191–6.
90. Daykin CA, Van Duynhoven JP, Groenewegen A, Dachtler M, Van Amelsvoort JM, Mulder TP. Nuclear magnetic resonance spectroscopic based studies of the metabolism of black tea polyphenols in humans. J Agric Food Chem 2005;53:1428–34.
91. Law WS, Huang PY, Ong ES, Ong CN, Li SF, Pasikanti KK, Chan EC. Metabonomics investigation of human urine after ingestion of green tea with gas chromatography/mass spectrometry, liquid chromatography/mass spectrometry and (1)H NMR spectroscopy. Rapid Commun Mass Spectrom 2008;22:2436–46.
92. van Velzen EJ, Westerhuis JA, van Duynhoven JP, van Dorsten FA, Grun CH, Jacobs DM, Duchateau GS, Vis DJ, Smilde AK.

The American Journal of Clinical Nutrition

Phenotyping tea consumers by nutrikinetic analysis of polyphenolic end-metabolites. J Proteome Res 2009;8:3317–30.
93. van Dorsten FA, Grun CH, van Velzen EJJ, Jacobs DM, Draijer R, van Duynhoven JPM. The metabolic fate of red wine and grape juice polyphenols in humans assessed by metabolomics. Mol Nutr Food Res 2010;54:897–908.
94. Jacobs DM, Fuhrmann JC, van Dorsten FA, Rein D, Peters S, van Velzen EJJ, Hollebrands B, Draijer R, van Duynhoven J, Garczarek U. Impact of short-term intake of red wine and grape polyphenol extract on the human metabolome. J Agric Food Chem 2012;60:3078–85.
95. Vázquez-Fresno R, Llorach R, Alcaro F, Rodriguez MA, Vinaixa M, Chiva-Blanch G, Estruch R, Correig X, Andres-Lacueva C. 1H-NMR-based metabolomic analysis of the effect of moderate wine consumption on subjects with cardiovascular risk factors. Electrophoresis 2012;33:2345–54.
96. Llorach R, Urpi-Sarda M, Jauregui O, Monagas M, Andres-Lacueva C. An LC-MS-based metabolomics approach for exploring urinary metabolome modifications after cocoa consumption. J Proteome Res 2009;8:5060–8.
97. Llorach R, Urpi-Sarda M, Tulipani S, Garcia-Aloy M, Monagas M, Andres-Lacueva C. Metabolomic fingerprint in patients at high risk of cardiovascular disease by cocoa intervention. Mol Nutr Food Res 2013;57:962–73.
98. Llorach R, Garrido I, Monagas MA, Urpi-Sarda M, Tulipani S, Bartolome BA, Andres-Lacueva C. Metabolomics study of human urinary metabolome modifications after intake of almond (Prunus dulcis (Mill.) D.A. Webb) skin polyphenols. J Proteome Res 2010;9:5859–67.
99. Rasmussen LG, Winning H, Savorani F, Ritz C, Engelsen SB, Astrup A, Larsen TM, Dragsted LO. Assessment of dietary exposure related to dietary GI and fibre intake in a nutritional metabolomic study of human urine. Genes Nutr 2012;7:281–93.
100. Johansson-Persson A, Barri T, Ulmius M, Onning G, Dragsted LO. LC-QTOF/MS metabolomic profiles in human plasma after a 5-week high dietary fiber intake. Anal Bioanal Chem 2013;405:4799–809.
101. Stanstrup J, Rasmussen J, Ritz C, Holmer-Jensen J, Hermansen K, Dragsted L. Intakes of whey protein hydrolysate and whole whey proteins are discriminated by LC-MS metabolomics. Metabolomics 2013;Nov 17 (Epub ahead of print; DOI 10.1007/s11306-013-0607-9).
102. Stella C, Beckwith-Hall B, Cloarec O, Holmes E, Lindon JC, Powell J, van der Ouderaa F, Bingham S, Cross AJ, Nicholson JK. Susceptibility of human metabolic phenotypes to dietary modulation. J Proteome Res 2006;5:2780–8.
103. Bertram HC, Hoppe C, Petersen BO, Duus JO, Molgaard C, Michaelsen KF. An NMR-based metabonomic investigation on effects of milk and meat protein diets given to 8-year-old boys. Br J Nutr 2007;97:758–63.
104. Xu J, Yang S, Cai S, Dong J, Li X, Chen Z. Identification of biochemical changes in lactovegetarian urine using 1H NMR spectroscopy and pattern recognition. Anal Bioanal Chem 2010;396:1451–63.
105. May DH, Navarro SL, Ruczinski I, Hogan J, Ogata Y, Schwarz Y, Levy L, Holzman T, McIntosh MW, Lampe JW. Metabolomic profiling of urine: response to a randomised, controlled feeding study of select fruits and vegetables, and application to an observational study. Br J Nutr 2013;110:1760–70.
106. Andersen MB, Rinnan A, Manach C, Poulsen SK, Pujos-Guillot E, Larsen TM, Astrup A, Dragsted LO. Untargeted metabolomics as a screening tool for estimating compliance to a dietary pattern. J Proteome Res 2014;13:1405–18.
107. Svensson BG, Akesson B, Nilsson A, Paulsson K. Urinary excretion of methylamines in men with varying intake of fish from the Baltic Sea. J Toxicol Environ Health 1994;41:411–20.
108. Koeth RA, Wang Z, Levison BS, Buffa JA, Org E, Sheehy BT, Britt EB, Fu X, Wu Y, Li L, et al. Intestinal microbiota metabolism of l-carnitine, a nutrient in red meat, promotes atherosclerosis. Nat Med 2013;19:576–85.
109. Altmaier E, Kastenmuller G, Romisch-Margl W, Thorand B, Weinberger KM, Illig T, Adamski J, Doring A, Suhre K. Questionnaire-based self-reported nutrition habits associate with serum metabolism as revealed by quantitative targeted metabolomics. Eur J Epidemiol 2011;26:145–56.
110. Sumner LW, Amberg A, Barrett D, Beale M, Beger R, Daykin C, Fan T, Fiehn O, Goodacre R, Griffin J, et al. Proposed minimum reporting standards for chemical analysis. Metabolomics 2007;3:211–21.
111. Theodoridis G, Gika HG, Wilson ID. LC-MS-based methodology for global metabolite profiling in metabonomics/metabolomics. Trends Analyt Chem 2008;27:251–60.
112. Scalbert A, Brennan L, Fiehn O, Hankemeier T, Kristal BS, van Ommen B, Pujos-Guyot E, Verheij E, Wishart D, Wopereis S. Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. Metabolomics 2009;5:435–58.
113. Issaq HJ, Van QN, Waybright TJ, Muschik GM, Veenstra TD. Analytical and statistical approaches to metabolomics research. J Sep Sci 2009;32:2183–99.
114. Beckmann M, Lloyd AJ, Haldar S, Seal C, Brandt K, Draper J. Hydroxylated phenylacetamides derived from bioactive benzoxazinoids are bioavailable in humans after habitual consumption of whole grain sourdough rye bread. Mol Nutr Food Res 2013;57:1859–73.
115. He X, Kozak M, Nimkar S. Ultra-sensitive measurements of 11-nor-delta(9)-tetrahydrocannabinol-9-carboxylic acid in oral fluid by microflow liquid chromatography-tandem mass spectrometry using a benchtop quadrupole/Orbitrap mass spectrometer. Anal Chem 2012;84:7643–7.
116. Barri T, Holmer-Jensen J, Hermansen K, Dragsted LO. Metabolic fingerprinting of high-fat plasma samples processed by centrifugation- and filtration-based protein precipitation delineates significant differences in metabolite information coverage. Anal Chim Acta 2012;718:47–57.
117. Gika HG, Theodoridis GA, Earll M, Snyder RW, Sumner SJ, Wilson ID. Does the mass spectrometer define the marker? A comparison of global metabolite profiling data generated simultaneously via UPLC-MS on two different mass spectrometers. Anal Chem 2010;82:8226–34.
118. Guo B, Chen B, Liu AM, Zhu WT, Yao SZ. Liquid chromatography-mass spectrometric multiple reaction monitoring-based strategies for expanding targeted profiling towards quantitative metabolomics. Curr Drug Metab 2012;13:1226–43.
119. Katajamaa M, Oresic M. Data processing for mass spectrometry-based metabolomics. J Chromatogr A 2007;1158:318–28.
120. Sugimoto M, Kawakami M, Robert M, Soga T, Tomita M. bioinformatics tools for mass spectroscopy-based metabolomic data processing and analysis. Current Bioinformatics 2012;7:96–108.
121. Savorani F, Kristensen M, Larsen FH, Astrup A, Engelsen SB. High throughput prediction of chylomicron triglycerides in human plasma by nuclear magnetic resonance and chemometrics. Nutr Metab (Lond) 2010;7:43.
122. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem 2006;78:779–87.
123. Pluskal T, Castillo S, Villar-Briones A, Oresic M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics 2010;11:395.
124. Lommen A, Kools HJ. MetAlign 3.0: performance enhancement by efficient use of advances in computer hardware. Metabolomics 2012;8:719–26.
125. Gürdeniz G, Kristensen M, Skov T, Dragsted LO. The effect of LC-MS data preprocessing methods on the selection of plasma biomarkers in fed vs. fasted rats. Metabolites 2012;2:77–99.
126. Kristensen M, Engelsen SB, Dragsted LO. LC-MS metabolomics top-down approach reveals new exposure and effect biomarkers of apple and apple-pectin intake. Metabolomics 2012;8:64–73.
127. Worley B, Halouska S, Powers R. Utilities for quantifying separation in PCA/PLS-DA scores plots. Anal Biochem 2013;433:102–4.
128. Acar E, Gurdeniz G, Rasmussen MA, Rago D, Dragsted LO, Bro R. Coupled matrix factorization with sparse factors to identify potential biomarkers in metabolomics. Int J Knowledge Discov Bioinformatics 2012;3:1–22.
129. Gürdeniz G, Hansen L, Rasmussen M, Acar E, Olsen A, Christensen J, Barri T, Tjønneland A, Dragsted L. Patterns of time since last meal revealed by sparse PCA in an observational LC-MS based metabolomics study. Metabolomics 2013;9:1073–81.
130. Hanhineva K, Barri T, Kolehmainen M, Pekkinen J, Pihlajamaki J, Vesterbacka A, Solano-Aguilar G, Mykkanen H, Dragsted LO, Urban JF Jr, et al. Comparative nontargeted profiling of metabolic changes in

tissues and biofluids in high-fat diet-fed Ossabaw pig. J Proteome Res 2013;12:3980–92.
131. Kjeldahl K, Bro R. Some common misunderstandings in chemometrics. J Chemometr 2010;24:558–64.
132. Wold S, Martens H, Wold H, Ruhe A, Kågstrøm B, eds. The multivariate calibration problem in chemistry solved by the PLS methods. In: Ruhe A, Kågstrøm B, eds. Proceedings of the Conference on Matrix Pencils, Lecture Notes in Mathematics. Heidelberg, Germany: Springer Verlag, 1983:286–93.
133. Smilde AK, Jansen JJ, Hoefsloot HCJ, Lamers R, van der Greef J, Timmerman ME. ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. Bioinformatics 2005;21:3043–8.
134. Rago D, Mette K, Gürdeniz G, Marini F, Poulsen M, Dragsted L. A LC-MS metabolomics approach to investigate the effect of raw apple intake in the rat plasma metabolome. Metabolomics 2013;9:1202–15.
135. van Velzen EJJ, Westerhuis JA, van Duynhoven JPM, van Dorsten FA, Hoefsloot HCJ, Jacobs DM, Smit S, Draijer R, Kroner CI, Smilde AK. Multilevel data analysis of a crossover designed human nutritional intervention study. J Proteome Res 2008;7:4483–91.
136. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Stat Med 1990;9:811–8.
137. Benjamini Y, Heller R, Yekutieli D. Selective inference in complex research. Philos Trans A Math Phys Eng Sci 2009;367:4255–71.
138. Xia J, Broadhurst D, Wilson M, Wishart D. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. Metabolomics 2013;9:280–99.
139. Del Rio D, Rodriguez-Mateos A, Spencer JPE, Tognolini M, Borges G, Crozier A. Dietary (poly)phenolics in human health: structures, bioavailability, and evidence of protective effects against chronic diseases. Antioxid Redox Signal 2013;18:1818–92.
140. van der Hooft JJJ, Vos RCH, Ridder L, Vervoort J, Bino RJ. Structural elucidation of low abundant metabolites in complex sample matrices. Metabolomics 2013;9:1009–18.
141. Wishart DS. Advances in metabolite identification. Bioanalysis 2011;3:1769–82.
142. Draper J, Enot DP, Parker D, Beckmann M, Snowdon S, Lin W, Zubair H. Metabolite signal identification in accurate mass metabolomics data with MZedDB, an interactive m/z annotation tool utilising predicted ionisation behaviour 'rules'. BMC Bioinformatics 2009;10:227.
143. Draper J, Lloyd A, Goodacre R, Beckmann M. Flow infusion electrospray ionisation mass spectrometry for high throughput, non-targeted metabolite fingerprinting: a review. Metabolomics 2013;9:4–29.
144. van der Hooft JJJ, Vervoort J, Bino RJ, Beekwilder J, de Vos RCH. Polyphenol identification based on systematic and robust high-resolution accurate mass spectrometry fragmentation. Anal Chem 2011;83:409–16.
145. Tikunov YM, Laptenok S, Hall RD, Bovy A, de Vos RCH. MSClust: a tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data. Metabolomics 2012;8:714–8.
146. Li L, Li R, Zhou J, Zuniga A, Stanislaus AE, Wu Y, Huan T, Zheng J, Shi Y, Wishart DS, et al. MyCompoundID: using an evidence-based metabolome library for metabolite identification. Anal Chem 2013;85:3401–8.
147. Gong P, Cui N, Wu L, Liang Y, Hao K, Xu X, Tang W, Wang G, Hao H. Chemicalome and metabolome matching approach to elucidating biological metabolic networks of complex mixtures. Anal Chem 2012;84:2995–3002.
148. Menikarachchi LC, Hill DW, Hamdalla MA, Mandoiu II, Grant DF. In silico enzymatic synthesis of a 400,000 compound biochemical database for non-targeted metabolomics. J Chem Inf Model 2013;53:2483–92.
149. Tohge T, Fernie AR. Web-based resources for mass-spectrometry-based metabolomics: a user's guide. Phytochemistry 2009;70:450–6.
150. Scalbert A, Andres-Lacueva C, Arita M, Kroon P, Manach C, Urpi-Sarda M, Wishart D. Databases on food phytochemicals and their health promoting effects. J Agric Food Chem 2011;59:4331–48.
151. American Chemical Society. SciFinder. Available from: https://scifinder.cas.org (cited 3 September 2013).
152. National Institutes of Standards and Technology. NIST 11 EPA/NIH Mass Spectral Database and Search Program (NIST11/NIST08). Available from: http://www.chemsw.com/Software-and-Solutions/Laboratory-Software/NIST-Mass-Spectral-Database/NIST-EPA-NIH-Mass-Spectral-Database.aspx (cited 3 September 2013).
153. Wolf S, Schmidt S, Muller-Hannemann M, Neumann S. In silico fragmentation for computer assisted identification of metabolite mass spectra. BMC Bioinformatics 2010;11:148.
154. Rojas-Cherto M, Peironcely JE, Kasper PT, van der Hooft JJJ, De Vos RCH, Vreeken RJ, Hankemeier T, Reijmers T. Metabolite identification using automated comparison of high resolution MSn spectral trees. Anal Chem 2012;84:5524–34.
155. HighChem. Mass Frontier. Available from: http://www.highchem.com/index.php/massfrontier (cited 3 September 2013).
156. Gerlich M, Neumann S. MetFusion: integration of compound identification strategies. J Mass Spectrom 2013;48:291–8.
157. Stanstrup J, Gerlich M, Dragsted LO, Neumann S. Metabolite profiling and beyond: approaches for the rapid processing and annotation of human blood serum mass spectrometry data. Anal Bioanal Chem 2013;405:5037–48.
158. Kuhl C, Tautenhahn R, Bottcher C, Larson TR, Neumann S. CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. Anal Chem 2012;84:283–9.
159. Leibniz Institute of Plant Biochemistry. CAMERA annotation of LC/MS data. Available from: http://msbi.ipb-halle.de/msbi/CAMERA (cited 3 September 2013).
160. Matsuda F, Nakabayashi R, Sawada Y, Suzuki M, Hirai MY, Kanaya S, Saito K. Mass spectra-based framework for automated structural elucidation of metabolome data to explore phytochemical diversity. Front Plant Sci 2011;2:40.
161. Stalmach A, Mullen W, Barron D, Uchida K, Yokota T, Cavin C, Steiling H, Williamson G, Crozier A. Metabolite profiling of hydroxycinnamate derivatives in plasma and urine after the ingestion of coffee by humans: identification of biomarkers of coffee consumption. Drug Metab Dispos 2009;37:1749–58.
162. Ridder L, van der Hooft JJJ, Verhoeven S, De Vos RCH, Bino RJ, Vervoort J. Automatic chemical structure annotation of an LC-MSn based metabolic profile from green tea. Anal Chem 2013;85:6033–40.
163. van der Hooft JJJ, de Vos RCH, Mihaleva V, Bino RJ, Ridder L, de Roo N, Jacobs DM, van Duynhoven JPM, Vervoort J. structural elucidation and quantification of phenolic conjugates present in human urine after tea intake. Anal Chem 2012;84:7263–71.
164. Kirchmair J, Howlett A, Peironcely JE, Murrell DS, Williamson MJ, Adams SE, Hankemeier T, van Buren L, Duchateau G, Klaffke W, et al. How do metabolites differ from their parent molecules and how are they excreted? J Chem Inf Model 2013;53:354–67.
165. Nicholson JK, Holmes E, Elliott P. The Metabolome-Wide Association Study: a new look at human disease risk factors. J Proteome Res 2008;7:3637–8.
166. Bictash M, Ebbels TM, Chan Q, Loo RL, Yap IKS, Brown IJ, de Iorio M, Daviglus ML, Holmes E, Stamler J, et al. Opening up the "black box": metabolic phenotyping and metabolome-wide association studies in epidemiology. J Clin Epidemiol 2010;63:970–9.
167. Rappaport SM. Biomarkers intersect with the exposome. Biomarkers 2012;17:483–9.
168. Saadatian-Elahi M, Slimani N, Chajes V, Jenab M, Goudable J, Biessy C, Ferrari P, Byrnes G, Autier P, Peeters PHM, et al. Plasma phospholipid fatty acid profiles and their association with food intakes: results from a cross-sectional study within the European Prospective Investigation into Cancer and Nutrition. Am J Clin Nutr 2009;89:331–46.
169. Chajès V, Jenab M, Romieu I, Ferrari P, Dahm CC, Overvad K, Egeberg R, Tjonneland A, Clavel-Chapelon F, Boutron-Ruault M-C, et al. Plasma phospholipid fatty acid concentrations and risk of gastric adenocarcinomas in the European Prospective Investigation into Cancer and Nutrition (EPIC-EURGAST). Am J Clin Nutr 2011;94:1304–13.
170. Nicholson JK, Holmes E, Wilson ID. Gut microorganisms, mammalian metabolism and personalized health care. Nat Rev Microbiol 2005;3:431–8.
171. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, DuGar B, Feldstein AE, Britt EB, Xiaoming F, Yoon-Mi C, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature 2011;472:57–63.
172. Tang WHW, Wang Z, Levison BS, Koeth RA, Britt EB, Fu X, Wu Y, Hazen SL. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. N Engl J Med 2013;368:1575–84.
173. Lee SM, Kwon GY, Kim K-O, Kim Y-S. Metabolomic approach for determination of key volatile compounds related to beef flavor in

glutathione-Maillard reaction products. Anal Chim Acta 2011;703:204–11.
174. Mena P, Calani L, Dall'Asta C, Galaverna G, Garcia-Viguera C, Bruni R, Crozier A, Del Rio D. Rapid and comprehensive evaluation of (poly)phenolic compounds in pomegranate (Punica granatum L.) juice by UHPLC-MSn. Molecules 2012;17:14821–40.
175. Moco S, Forshed J, De Vos RCH, Bino RJ, Vervoort J. Intra- and inter-metabolite correlation spectroscopy of tomato metabolomics data obtained by liquid chromatography-mass spectrometry and nuclear magnetic resonance. Metabolomics 2008;4:202–15.
176. Adhikari KB, Laursen BB, Laerke HN, Fomsgaard IS. bioactive benzoxazinoids in rye bread are absorbed and metabolized in pigs. J Agric Food Chem 2012;60:2497–506.
177. Morand C, Dubray C, Milenkovic D, Lioger D, Martin JF, Scalbert A, Mazur A. Hesperidin contributes to the vascular protective effects of orange juice: a randomized crossover study in healthy volunteers. Am J Clin Nutr 2011;93:73–80.
178. Lan K, Xie G, Jia W. Towards polypharmacokinetics: pharmacokinetics of multicomponent drugs and herbal medicines using a metabolomics approach. Evid Based Complement Alternat Med 2013;2013:819147.
179. Johnson CH, Slanar O, Krausz KW, Kang DW, Patterson AD, Kim J-H, Luecke H, Gonzalez FJ, Idle JR. Novel metabolites and roles for alpha-tocopherol in humans and mice discovered by mass spectrometry-based metabolomics. Am J Clin Nutr 2012;96:818–30.
180. Suhre K, Shin S-Y, Petersen A-K, Mohney RP, Meredith D, Waegele B, Altmaier E, Deloukas P, Erdmann J, Grundberg E, et al. Human metabolic individuality in biomedical and pharmaceutical research. Nature 2011;477:54–60.
181. Illig T, Gieger C, Zhai GJ, Romisch-Margl W, Wang-Sattler R, Prehn C, Altmaier E, Kastenmuller G, Kato BS, Mewes HW, et al. A genome-wide perspective of genetic variation in human metabolism. Nat Genet 2010;42:137–41.
182. Kettunen J, Tukiainen T, Sarin A-P, Ortega-Alonso A, Tikkanen E, Lyytikainen L-P, Kangas AJ, Soininen P, Wuertz P, Silander K, et al. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. Nat Genet 2012;44:269–76.
183. Lampe JW. Interindividual differences in response to plant-based diets: implications for cancer risk. Am J Clin Nutr 2009;89:1553S–7S.
184. Reszka E, Wasowicz W, Gromadzinska J. Genetic polymorphism of xenobiotic metabolising enzymes, diet and cancer susceptibility. Br J Nutr 2006;96:609–19.
185. Willett WC. Balancing life-style and genomics research for disease prevention. Science 2002;296:695–8.
186. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 2003;32:1–22.
187. Schroeter H, Heiss C, Balzer J, Kleinbongard P, Keen CL, Hollenberg NK, Sies H, Kwik-Uribe C, Schmitz HH, Kelm M. (-)-Epicatechin mediates beneficial effects of flavanol-rich cocoa on vascular function in humans. Proc Natl Acad Sci USA 2006;103:1024–9.
188. Fiehn O, Robertson D, Griffin J, van der Werf M, Nikolau B, Morrison N, Sumner L, Goodacre R, Hardy N, Taylor C, et al. The Metabolomics Standards Initiative (MSI). Metabolomics 2007;3:175–8.
189. Metabolomics Standards Initiative. Homepage. Available from: http://msi-workgroups.sourceforge.net/ (cited 3 September 2013).
190. Metabolomics Forum. Homepage. Available from: http://metabolomics-forum.com (cited 3 September 2013).
191. NIH. ClinicalTrials.gov. Available from: www.ClinicalTrials.gov (cited 3 September 2013).
192. T'Jollyn H, Boussery K, Mortishire-Smith RJ, Coe K, De Boeck B, Van Bocxlaer JF, Mannens G. Evaluation of three state-of-the-art metabolite prediction software packages (Meteor, MetaSite, and StarDrop) through independent and synergistic use. Drug Metab Dispos 2011;39:2066–75.
193. Piechota P, Cronin MTD, Hewitt M, Madden JC. pragmatic approaches to using computational methods to predict xenobiotic metabolism. J Chem Inf Model 2013;53:1282–93.
194. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2013;41:D36–42.
195. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res 2013;41:D991–5.
196. Riffle M, Eng JK. Proteomics data repositories. Proteomics 2009;9:4653–63.
197. Haug K, Salek RM, Conesa P, Hastings J, de Matos P, Rijnbeek M, Mahendraker T, Williams M, Neumann S, Rocca-Serra P, et al. MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data. Nucleic Acids Res 2013;41:D781–6.
198. van Ommen B, Bouwman J, Dragsted LO, Drevon CA, Elliott R, de Groot P, Kaput J, Mathers JC, Muller M, Pepping F, et al. Challenges of molecular nutrition research 6: the nutritional phenotype database to store, share and evaluate nutritional systems biology studies. Genes Nutr 2010;5:189–203.