# Polyphenol metabolome in human urine and its association with intake of polyphenol-rich foods across European countries[1,2]

*William MB Edmands,[3,19] Pietro Ferrari,[3] Joseph A Rothwell,[3] Sabina Rinaldi,[3] Nadia Slimani,[3] Dinesh K Barupal,[3] Carine Biessy,[3] Mazda Jenab,[3] Françoise Clavel-Chapelon,[4–6] Guy Fagherazzi,[4–6] Marie-Christine Boutron-Ruault,[4–6] Verena A Katzke,[7] Tilman Kühn,[7] Heiner Boeing,[8] Antonia Trichopoulou,[9–11] Pagona Lagiou,[10–12] Dimitrios Trichopoulos,[9,11,12] Domenico Palli,[13] Sara Grioni,[14] Rosario Tumino,[15] Paolo Vineis,[16,17] Amalia Mattiello,[18] Isabelle Romieu,[3] and Augustin Scalbert[3]**

[3]International Agency for Research on Cancer, Lyon, France; [4]French Institute of Health and Medical Research (Inserm), Centre for Research in Epidemiology and Population Health, U1018, Nutrition, Hormones and Women's Health Team, Villejuif, France; [5]Université Paris Sud, UMRS 1018, Villejuif, France; [6]Institut Gustave Roussy, Villejuif, France; [7]Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany; [8]German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany; [9]Hellenic Health Foundation, Athens, Greece; [10]Department of Hygiene, Epidemiology and Medical Statistics, University of Athens Medical School, Athens, Greece; [11]Bureau of Epidemiologic Research, Academy of Athens, Athens, Greece; [12]Department of Epidemiology, Harvard School of Public Health, Boston, MA; [13]Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute, Florence, Italy; [14]Epidemiology and Prevention Unit, Fondazione IRCCS, Istituto Nazionale dei Tumori, Milan, Italy; [15]Cancer Registry and Histopathology Unit, "Civic - M.P. Arezzo" Hospital, Provincial Health Unit Ragusa, Italy; [16]Medical Research Council, Public Health England Center for Environment and Health, School of Public Health, Imperial College London, London, United Kingdom; [17]HuGeF Foundation, Turin, Italy; and [18]Department of Clinical Medicine and Surgery, Federico II University, Naples, Italy

## ABSTRACT

**Background:** An improved understanding of the contribution of the diet to health and disease risks requires accurate assessments of dietary exposure in nutritional epidemiologic studies. The use of dietary biomarkers may improve the accuracy of estimates.

**Objective:** We applied a metabolomic approach in a large cohort study to identify novel biomarkers of intake for a selection of polyphenol-containing foods. The large chemical diversity of polyphenols and their wide distribution over many foods make them ideal biomarker candidates for such foods.

**Design:** Metabolic profiles were measured with the use of high-resolution mass spectrometry in 24-h urine samples from 481 subjects from the large European Prospective Investigation on Cancer and Nutrition cohort. Peak intensities were correlated to acute and habitual dietary intakes of 6 polyphenol-rich foods (coffee, tea, red wine, citrus fruit, apples and pears, and chocolate products) measured with the use of 24-h dietary recalls and food-frequency questionnaires, respectively.

**Results:** Correlation ($r > 0.3$, $P < 0.01$ after correction for multiple testing) and discriminant [pcorr (1) $> 0.3$, VIP $> 1.5$] analyses showed that >2000 mass spectral features from urine metabolic profiles were significantly associated with the consumption of the 6 selected foods. More than 80 polyphenol metabolites associated with the consumption of the selected foods could be identified, and large differences in their concentrations reflecting individual food intakes were observed within and between 4 European countries. Receiver operating characteristic curves showed that 5 polyphenol metabolites, which are characteristic of 5 of the 6 selected foods, had a high predicting ability of food intake.

**Conclusion:** Highly diverse food-derived metabolites (the so-called food metabolome) can be characterized in human biospecimens through this powerful metabolomic approach and screened to identify novel biomarkers for dietary exposures, which are ultimately essential to better understand the role of the diet in the cause of chronic diseases. *Am J Clin Nutr* 2015;102:905–13.

**Keywords:** dietary biomarkers, food metabolome, polyphenols, flavonoids, phenolic acids, coffee, tea, red wine, citrus fruits, EPIC

## INTRODUCTION

The human organism is constantly exposed to diverse environmental chemicals, either natural or man-made, that are present in food and drinking water, air, or any drug or consumer products. In particular, >27,000 compounds have been described in foods (1, 2), and many of them can be absorbed in the gut and metabolized in tissues or by the gut microbiota. These food-derived metabolites constitute the so-called food metabolome (3). Some of these metabolites have been used as dietary biomarkers for monitoring exposures to specific components of the diet in populations and for

[1] Supported by the European Union (NutriTech FP7-KBBE-2011-5 grant 289511, EUROCAN FP7-KBBE-2010.2.4.1-2 grant 60791), the European Commission (Directorate General for Health and Consumer Affairs), and the International Agency for Research on Cancer. The national cohorts are supported by 3M, the Mutuelle Générale de l'Education Nationale, and the Institut National de la Santé et de la Recherche Medicale (France); Deutsche Krebshilfe, Deutsches Krebsforschungszentrum, and the Federal Ministry of Education and Research (Germany); the Ministry of Health and Social Solidarity, the Stavros Niarchos Foundation, and the Hellenic Health Foundation (Greece); and the Italian Association for Research on Cancer and the National Research Council (Italy).

[2] Supplemental Figure 1, Supplemental Subjects and Methods, and Supplemental Tables 1–12 are available from the "Supplemental data" link in the online posting of the article and from the same link in the online table of contents at http://ajcn.nutrition.org.

[19] Present address: School of Public Health, University of California, Berkeley, CA.

*To whom correspondence should be addressed. E-mail: scalberta@iarc.fr.

Received October 24, 2014. Accepted for publication July 21, 2015.

First published online August 12, 2015; doi: 10.3945/ajcn.114.101881.

studying associations between dietary exposures and disease risk (3–5).

Thus far, ≤100 dietary biomarkers have been measured in various cohort studies, and these biomarkers represent only a minor fraction of the food metabolome (3, 6). More biomarkers remain to be identified, and metabolomic approaches have already been successfully used to identify novel dietary biomarkers such as proline betaine and *S*-methyl-L-cysteine sulfoxide for citrus fruit and cruciferous vegetable intakes, respectively (7–11). Both nuclear magnetic resonance spectroscopy and mass spectrometry (MS)[20] have been used in these metabolomic studies. However, the high sensitivity of modern high-resolution MS allows for the measurement of thousands of metabolites in small human biospecimens (12, 13). This technical progress has opened new perspectives to measure the food metabolome in a more-comprehensive way and to provide a richer view of dietary exposures (3).

In addition, the measurement of the food metabolome may improve the accuracy of the dietary assessment. Usual approaches to measure dietary exposure in large epidemiologic studies are largely based on self-administered food-frequency questionnaires. The resulting dietary measurements are prone to random and systematic measurement errors, which are, in turn, a cause of bias in the evaluation of associations between diet and risk of diseases (14). In contrast, dietary biomarkers are objective measures that can either be directly used in etiological models or can help disclose the measurement-error structure in dietary assessments (15, 16), thereby improving the accuracy of estimates. In addition, a more-comprehensive measurement of the food metabolome should permit the discovery of unexpected novel risk factors for diseases as exemplified by the identification of choline and its microbial products as risk factors for cardiovascular diseases (17).

In this article, we compared the polyphenol metabolome, which is a major fraction of the food metabolome, in urine samples from 481 free-living subjects from 4 European countries as part of the EPIC (European Prospective Investigation on Cancer and Nutrition) cohort. Polyphenols were selected for their high abundance in the diet, large chemical diversity, wide distribution over a large diversity of foods, their high prospect as dietary biomarkers, and the high interest in their biological and health properties (18–21). Data from urine metabolic profiles were mined to identify good predictors of intake for 6 polyphenol-rich foods that may improve measurements of intake in epidemiologic studies.

## METHODS

### Subject selection, dietary data, and urine samples

The EPIC study was designed to investigate the relations between diet, nutritional status, and lifestyle and environmental factors and the incidence of cancer and other chronic diseases. The EPIC is among the largest studies on diet and cancer with more than than one-half million (520,000) participants recruited between 1992 and 2000 in 23 centers in 10 European countries (22). For this study, participants were selected from a subset ($n$ = 1072) of the EPIC cohort for whom an archived 24-h urine collection and a single standardized 24-h dietary recall (24-HDR) were taken on the same day (23) and a dietary food-frequency questionnaire per subject was available. The standardized 24-HDR and dietary questionnaires used in each participating country have been extensively validated within the EPIC study (15, 23, 24). The collection of urine samples started just after the first pass of the day at 0700 and ended at the same time the next day with the inclusion of the first urine pass. For collection, subjects were given two 2-L containers, each of which included 2 g boric acid as preservative. Dietary intake recorded in the 24-HDR included all foods ingested during the same period. The completeness of the collection of 24-h urine samples was monitored with the use of *p*-aminobenzoic acid that was given to participants in a tablet form. Samples with a *p*-aminobenzoic acid recovery <85% or >110% were excluded from the study (23). A total of 481 urine samples were finally selected for the study. Urine samples were collected between 1995 and 1999 and stored at −20°C. The following 6 food items were selected: coffee, red wine, citrus fruit (including orange, mandarin, lemon, grapefruit, and lime), tea, apples and pears, and chocolate products (from chocolate bars, candy bars, and paste). The selection of food items was based on their high contents in polyphenols and possible role in the prevention of various diseases such as cardiovascular diseases, diabetes, or cancers (25–29). A summary of the consumption of the 6 polyphenol-rich foods and their country-specific patterns in the 481 subjects is shown in **Supplemental Tables 1** and **2**.

Informed consent was provided by each participant in the study. All participants gave consent for future analyses of their urine samples, and the Ethics Committee of the International Agency for Research on Cancer approved the metabolomic analyses.

### Metabolomic profiling with the use of MS

Urine concentrations were first normalized before data acquisition to remove any unwanted variance of volume by dilution with water on the basis of the specific gravity measured with the use of refractometry (30). Diluted urine samples, blanks, and pooled urine quality-control (QC) samples were analyzed with the use of a ultra-high performance liquid chromatography system (Infinity 1290, Agilent Technologies Inc.) coupled to a quadrupole time-of-flight (Q-TOF) mass spectrometer (6550, Agilent Technologies Inc., France). The UPLC column (ACQUITY HSS T3, 1.8 $\mu$m, 2.1 mm × 100mm; Waters, Saint-Quentin en Yvelines) was held at 50°C, with a mobile-phase flow rate of 0.36 mL/min. Urine samples were maintained at 4°C in a thermostated autosampler before injection (1 $\mu$L). A 10-min linear gradient was applied with solvent A, which was composed of 0.1% formic acid in water, and solvent B, which was composed of 0.1% formic acid in methanol. The following gradient was used: 2% B for 2 min was ramped successively to 30% B in 6 min, 100% B in 3 min, and finally maintained at 100% B for one additional minute. The Q-TOF was operated in negative electrospray ionization mode from 50 to 1100 $m/z$. Pooled QC samples were injected every 10th injection. See **Supplemental Subjects and Methods** for details.

[20] Abbreviations used: EPIC, European Prospective Investigation on Cancer and Nutrition; MS, mass spectrometry; MS/MS, tandem mass spectrometry; O-PLS-DA, orthogonal partial least-squares discriminant analysis; ppm, parts per million; QC, quality control; ROC, receiver operator characteristic; 24-HDR, 24-h dietary recall.

### Raw data processing and data analysis

MS raw data (.d) files were converted to the .mzxml cross-platform open file format before processing with the use of the XCMS platform (version 1.36.0) for nonlinear peak alignment (31) and the fully automated software MetMSLine (32) for automated zero-peak filling, generalized-log transformation, correction for signal drift across the analytic batch with the use of locally weighted scatterplot smoothing ,and outlier removal (see Supplemental Subjects and Methods for more details). Of the 481 subjects initially included in the study, a total of 476 subjects were retained, and 5 subjects were excluded because of outlying urinary metabolic profiles. Mass spectral features were filtered by a <30% relative SD cutoff within the repeated pooled QC injections.

A large list of dietary biomarkers were initially identified by computing Pearson correlation coefficients between peak areas of all MS features and 24-HDR measurements of each food item for each of the 476 subjects. Nonzero dietary measurements were included in the analysis. For each correlation, *P* values were calculated and corrected for the false-discovery rate with the use of the Benjamini-Hochberg method (33) to account for the large number of comparisons performed. Partial parametric Pearson correlation coefficients were calculated with the use of the variance-covariance matrix method. Statistical significance was assessed at the 1% level. Because a large number of MS features were identified, correlation coefficients >0.3 were further retained in the analysis.

In a second phase and for all the same food items, dietary biomarkers were also identified with the use of a pairwise orthogonal partial least-squares discriminant analysis (O-PLS-DA) whereby the top and bottom quintiles of the 24-HDR of each polyphenol-rich food were compared. Models were calculated on preprocessed, outlier-filtered data, which were centered and Pareto scaled in the SIMCA-P+ program (version 13; Umetrics). Biomarkers were identified on the basis of 4 O-PLS-DA model filtration criteria as follows: the magnitude of covariance [p(1) > 0.018], minimum O-PLS-DA loadings coefficient [pcorr(1) > 0.3], a variable influence on projection (VIP) >1.5, and the variable influence on projection after the subtraction of the jack-knifed 95% CI (>0.1) (32).

Correlations between mass spectral features that met the O-PLS-DA biomarker identification criteria were determined by the calculation of a Pearson correlation coefficient matrix. This interfeature correlation matrix was hierarchically clustered with the use of the average linkage method with dissimilarity calculated by the correlation metric (one-correlation coefficient) and finally visualized as a heat map with the use of the gplots package in R software (version 3.1.1, 64-bit).

O-PLS-DA models were also calculated to compare metabolic profiles between countries. Receiver operator characteristic (ROC) curves were calculated with the use of the pROC package in R software by modeling top and bottom 24-HDR quintiles of each food item in turn. A nonparametric bootstrap resampling was performed ($n$ = 1000 iterations) for the ROC analysis of every polyphenol metabolite (**Supplemental Table 3**). A 2-sided, nonparametric 95% CI was calculated with the use of the percentiles of the bootstrap distribution. Calculations were performed in the R program (version 3.1.1 64-bit) with the use of the pROC (version 1.7.3) and boot (version 1.3–13) packages.

### Metabolite annotation

Chemical adducts were annotated automatically on the basis of high mass accuracy tolerance [<10 parts per million (ppm)] from a list of potential mass shifts that are commonly seen in liquid chromatography–MS data that result from the physical process of electrospray ionization (32). Unknown biomarkers were automatically annotated (32) with the use of mono-isotopic mass matching (<10 ppm) with biologically plausible metabolites of the Phenol-Explorer database (http://phenol-explorer.eu/) (34) and all possible phase II metabolites (sulfate esters, glucuronides, and *N*-acetylcysteine conjugates). The chemical identity was confirmed by comparing tandem mass spectrometry (MS/MS) fragmentation spectra obtained with the use of a novel data-dependent liquid chromatography–MS/MS method that was based on the sequential iterative exclusion of MS features for which an MS/MS spectrum has already been acquired in the previous round (see Supplemental Subjects and Methods for details). Fragmentation experiments were conducted with the use of a sample that was prepared by pooling individual samples from highest consumers of the 6 foods of interest ($n$ = 19) and also the pooled QC ($n$ = 481). Urinary fragmentation spectra were compared with those extracted from the literature and online open-access MS/MS databases or obtained from authentic chemical standards when available (**Supplemental Tables 4–9**). A short list of 34 metabolites characterized by their high correlation level with intakes of the 6 foods and highest evidence supporting their assignment (MS/MS match and availability of chemical standards) is given in **Table 1**.

## RESULTS

### Large number of features in the urinary metabolome correlate with food intake

A total of 481 subjects from the EPIC cohort took part in the study; 59% were women and 41% were men with a mean ± SD age of 55.3 ± 8.4 y at the time of the collection of urine samples and dietary intake data. Study subjects were from France (Ile de France), Germany (Heidelberg and Potsdam), Greece (nation wide), and Italy (Florence, Naples, Ragusa, Turin, and Varese) and were selected for the availability of both a 24-h urine sample and 24-HDR collected on the same day. The metabolome was measured with the use of high-resolution MS, and 14,323 MS features (i.e., detected ions) that passed the QC test were semiquantified in the urine samples (**Supplemental Figure 1**).

This data set was mined to identify MS features correlated to the acute intake of the following 6 foods that are rich in polyphenols: coffee, tea, red wine, citrus fruit, apples and pears, and chocolate products. With the use of a significance level of 1%, Benjamini-Hochberg–adjusted *P* values (35), and a minimum threshold value of 0.3 for the correlation coefficient, a total of 1064, 107, 829, 317, 19, and 2 MS features ($n$ = 2272) were shown to correlate with 24-HDR measurements of coffee, tea, red wine, citrus fruit, apples and pears, and chocolate products, respectively (**Figure 1**A, Supplemental Figure 1).

Given the low number of MS features associated with intakes of apples and pears and chocolate products, we also tested the O-PLS-DA to relate MS features to the top ($n$ = 95) and bottom ($n$ = 95) 24-HDR quintiles, thereby comparing, for each food, the most-extreme consumers according to their acute food intakes. This

**TABLE 1**
Food-derived metabolites in urine of free-living subjects from the EPIC cohort correlated with acute intakes of 6 polyphenol-rich foods[1]

| ID | Assignment[2] | Food | VIP | ROC AUC, % |
|---|---|---|---|---|
| 1 | Dihydroferulic acid sulfate (I) | Coffee | 3.92 | 95.5 |
| 2 | Guaiacol glucuronide | Coffee | 3.73 | 95.3 |
| 3 | Feruloylquinic acid (I) | Coffee | 3.41 | 91.9 |
| 4 | Ferulic acid sulfate (I) | Coffee | 3.77 | 95.2 |
| 5 | Feruloylquinic acid glucuronide (I) | Coffee | 3.77 | 94.3 |
| 6 | 3-*O*-Caffeoylquinic acid (I) | Coffee | 2.89 | 88.4 |
| 7 | *p*-Coumaric acid sulfate | Coffee | 3.38 | 91 |
| 8 | Caffeic acid sulfate (I) | Coffee | 3.04 | 91.4 |
| 9 | Ferulic acid glucuronide (I) | Coffee | 2.9 | 89.3 |
| 10 | Hydroxyhippuric acid (I) | Coffee | 2.24 | 86.1 |
| 11 | Dihydrocaffeic acid sulfate | Coffee | 2.63 | 85 |
| 12 | *m*-Coumaric acid sulfate | Coffee | 2.28 | 79.9 |
| | | Red wine | 0.52 | 52.7 |
| 13 | Dihydroferulic acid glucuronide (I) | Coffee | 2.63 | 85.7 |
| 14 | *p*-Hydroxyphenyllactic acid | Coffee | 2.06 | 83.2 |
| 15 | Guaiacol sulfate | Coffee | 2.23 | 86 |
| 16 | Ethylcatechol glucuronide | Coffee | 2.82 | 88 |
| 17 | Gallic acid ethyl ester sulfate | Red wine | 6.49 | 91.9 |
| 18 | Hydroxytyrosol sulfate | Red wine | 3.2 | 76.7 |
| 19 | Dihydroresveratrol glucuronide | Red wine | 4.9 | 86.9 |
| 20 | Syringic acid sulfate | Red wine | 2.64 | 72.8 |
| 21 | Naringenin glucuronide | Citrus fruit | 6.23 | 91.3 |
| 22 | Hesperetin glucuronide sulfate | Citrus fruit | 5.55 | 86.8 |
| 23 | Hesperetin glucuronide (I) | Citrus fruit | 6.25 | 89 |
| 24 | Methylgallic acid sulfate (I) | Tea | 6.62 | 81.6 |
| | | Red wine | 4.61 | 83.6 |
| 25 | 4-*O*-Methylgallic acid | Tea | 6.81 | 83.9 |
| | | Red wine | 4.83 | 84.7 |
| 26 | Dihydroxyphenyl-γ-valerolactone sulfate | Tea | 6.58 | 79.3 |
| 27 | Pyrogallol sulfate (I) | Tea | 3.59 | 70.4 |
| 28 | Hydroxyphenylvaleric acid glucuronide | Tea | 3.46 | 65.3 |
| 29 | Methyl(epi)catechin sulfate (I) | Tea | 3.01 | 66.9 |
| | | Apples and pears | 4.17 | 70.7 |
| | | Chocolate | 4.93 | 74.4 |
| 30 | Phloretin glucuronide | Apples and pears | 4.58 | 75.8 |
| 31 | Dihydroxyphenyl-γ-valerolactone sulfate | Apples and pears | 2.91 | 63.2 |
| 32 | 4-Hydroxy-(3′,4’-dihydroxyphenyl)valeric acid sulfate | Chocolate | 2.96 | 66.4 |
| 33 | Dihydroxyphenyl-γ-valerolactone glucuronide | Chocolate | 2.24 | 63.6 |
| | | Tea | 2.94 | 59.2 |
| 34 | Vanillic acid sulfate | Chocolate | 2.42 | 62.1 |

[1]VIP is a variable that summarizes the importance of *X* variables to the O-PLS-DA model. Variables with values >1.5 were the most influential in the model. The ROC AUC is a measure of the sensitivity and specificity of the biomarker for a food. EPIC, European Prospective Investigation on Cancer and Nutrition; ID, identification number; O-PLS-DA, orthogonal partial least-squares discriminant analysis; ROC, receiver operator characteristic; VIP, variable influence in projection.

[2]Number in parentheses refers to the isomers when several isomers were detected but not fully resolved (see Supplemental Tables 4–9).

approach led to the identification of 1229, 537, 566, 131, 55, and 306 MS features that were most discriminant for coffee, red wine, citrus fruit, tea, apples and pears, and chocolate products, respectively (totaling 2824 MS features for the 6 foods). A description of results from O-PLS-DA models is summarized in **Supplemental Table 10**. The O-PLS-DA allowed for the identification of a higher number of biomarkers than did correlation analyses (except for red wine), and these biomarkers included all or most of those identified through correlation analyses, which showed the large overlap of the 2 methods. A Pearson correlation coefficient matrix was calculated for the 2824 markers. An unsupervised hierarchical clustering analysis of these markers combined with retention-time matching (±2 s) led to the identification of 374 retention time clusters that accounted for 64% of the significant features, which, subsequently, greatly facilitated the identification (±10 ppm) of isotopomers, chemical adducts, and fragments (25% of all significant features) formed in the mass spectrometer source (Supplemental Subjects and Methods).

### Concentrations of polyphenol metabolites in urine correlate with both acute and habitual intakes of polyphenol-rich foods

To annotate the metabolites associated with intakes of the 6 selected foods, the 2824 MS features previously identified were screened and matched on the basis of their accurate monoisotopic

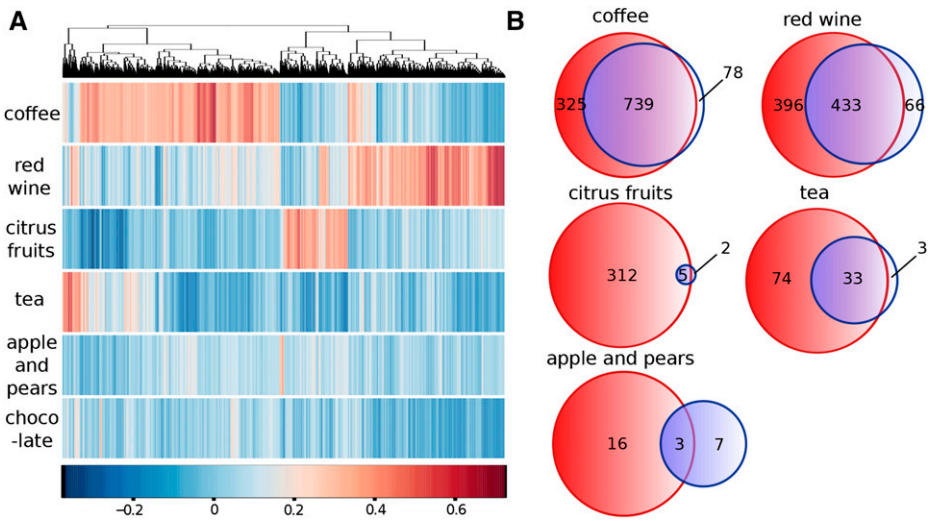


**FIGURE 1** Mass spectral features observed in urine samples from the European Prospective Investigation on Cancer and Nutrition cohort showing significant correlations with intakes of 6 polyphenol-rich foods. (A) Hierarchical clustering analysis of the 2272 mass spectral features linearly associated with acute intakes of the 6 polyphenol-rich foods. The color scale indicates the level of correlation (24-h dietary recalls; Pearson correlation coefficient >0.3; *F* test with Benjamini-Hochberg correction: $P < 0.01$) and reveal clusters of mass spectral features that are characteristic of each food. (B) Venn diagrams visualizing the overlap between mass spectral features correlated with acute 24-HDRs (red circles) and habitual food intakes (with the use of FFQs; blue circles) of 5 of 6 selected foods. Numbers of acute and habitual consumers (24-HDR, FFQ) were as follows: coffee ($n = 413, 445$), red wine ($n = 120, 128$), citrus fruit ($n = 185, 460$), tea ($n = 119, 313$), and apples and pears ($n = 228, 460$). The Venn diagram for chocolate products is not shown because only 2 mass spectral features were demonstrated to be significantly correlated to its acute intake. A broad overlap can be seen between biomarkers of acute and habitual intakes for some foods, particularly for coffee, tea, or red wine, which are more-regularly consumed in the populations considered. FFQ, food-frequency questionnaire; 24-HDR, 24-h dietary recall.

mass with those expected from known polyphenol metabolites. The Phenol-Explorer database was used to generate a list of 123 polyphenol entries made of all polyphenols known in the selected foods and of all metabolites known to be formed from these polyphenols or their food sources (34, 36). In addition, all possible combinations of phase II conjugates were also calculated for each database entry and added to the list of expected mass spectral features. All expected polyphenol metabolites were queried against the unknown MS features associated with the consumption of the selected foods, and 526 hits (<10-ppm mass accuracy) were identified. MS/MS fragmentation spectra were acquired for a fraction (24%) of these tentative assignments. Fragments and neutral losses were automatically annotated and queried against MS/MS spectra in on-line databases such as the Human Metabolome Database, Metlin, Massbank, and the ReSpect database. All assignments of polyphenol metabolites ($n = 83$) showed a strong biological plausibility on the basis of previous literature reports and were confirmed with authentic standards when available (see Table 1 and Supplemental Tables 4–9 for a complete list of annotated metabolites).

ROC curves were calculated and AUCs and associated 95% CIs were used to compare for each of the 83 polyphenol metabolites to evaluate their capacities to predict intakes of the 6 polyphenol-rich foods (Table 1, **Figure 2**). Compounds with the greatest predictive ability were dihydroferulic acid sulfate for coffee (AUC: 95.5%), gallic acid ethyl ester for red wine (AUC: 91.9%), naringenin glucuronide for citrus fruit (AUC: 91.3%), 4-*O*-methylgallic acid for tea (AUC: 83.9%), phloretin glucuronide for apples and pears (AUC: 75.8%), and methyl(epi)catechin sulfate for chocolate products (AUC: 74.4%) (Figure 2A, B).

Four of these compounds (dihydroferulic acid sulfate, gallic acid ethyl ester, naringenin glucuronide, and phloretin glucuronide) are known to be derived from precursors that originate predominantly or exclusively from the associated foods (**Supplemental Table 11**), which supports their possible use as biomarkers of food intake. The 2 remaining compounds were shown to be also associated with intakes of other foods as follows: 4-*O*-methylgallic acid with red wine (AUC: 84.7%), and methyl(epi)catechin sulfate with apples and pears (AUC: 70.7%), and tea (AUC: 74.4%) (Table 1).

We first compared the correlation between 4-*O*-methylgallic acid and tea after adjustment for red wine. Unadjusted and adjusted correlation values were 0.55 and 0.63, respectively, which suggested a limited confounding effect of red wine in the association between tea and 4-*O*-methylgallic acid. However, similar correlations were observed between 4-*O*-methylgallic acid and red wine with and without adjustment for tea intake, which indicated a lack of specificity in the use of 4-*O*-methylgallic acid as a marker of tea (and red wine) intake.

Similarly, the formation of methyl(epi)catechin sulfate from 4 different precursors (catechin, epicatechin, and their *O*-galloyl-esters) abundant in foods such as tea, wine, and apples (Supplemental Table 11) limits their possible use as biomarkers of intake for chocolate products. A number of other polyphenol metabolites were also shown to be associated with the intake of chocolate products with ROC AUC values that ranged from 62.1% to 66.4% (Table 1, Figure 2C). Two polyphenol metabolites [i.e., 4-hydroxy-(3′,4′-dihydroxyphenyl)valeric acid sulfate and dihydroxyphenyl-γ-valerolactone glucuronide) are also metabolites of (epi)catechin, which is a polyphenol that is present in tea, wine, and apples, that may act as confounders. Vanillic acid sulfate was also associated with the intake of chocolate products (Table 1, assignment 34) and is derived from vanillin, which is a common ingredient in many chocolate products. However, vanillin is also used as an additive in other food products, and its specificity for chocolate remains to be established.

Besides the possible consumption of other foods that contain the same polyphenol precursors, other confounders may also limit the accuracy of polyphenol biomarkers used to assess intakes

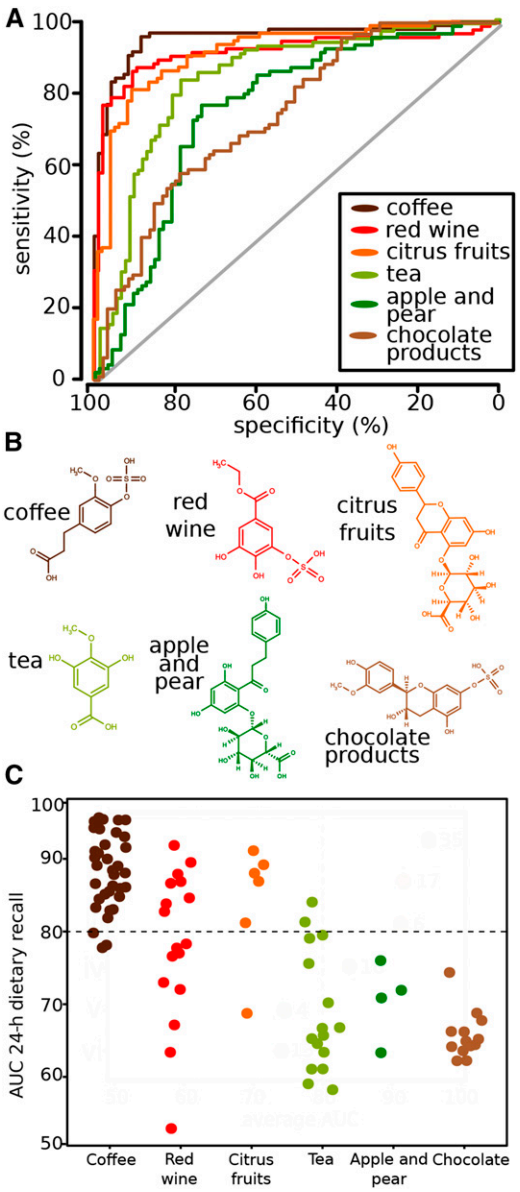


**FIGURE 2** Polyphenol metabolites as predictor of food consumption (24-h dietary recalls). (A) Receiver operator characteristic curves of the most-highly predictive polyphenol metabolites of the 6 polyphenol-rich foods. (B) Chemical structures of the following most-highly predictive metabolites: dihydroferulic acid sulfate for coffee (AUC: 95.5%), gallic acid ethyl ester for red wine (AUC: 91.9%), naringenin glucuronide for citrus fruit (AUC: 91.3%), 4-*O*-methylgallic acid (AUC: 83.9%) for tea, phloretin glucuronide for apples and pears (AUC: 75.8%), and methyl(epi)catechin sulfate for chocolate products (AUC: 74.4%). (C) Capacity of all polyphenol metabolites identified to predict intakes of the following 6 polyphenol-rich foods: coffee ($n$ = 35), red wine ($n$ = 17), citrus fruit ($n$ = 6), tea ($n$ = 16), apples and pears ($n$ = 4), and chocolate products ($n$ = 13).

of polyphenol-containing foods. More particularly, amounts of polyphenol biomarkers result from the action of transporters and phase I and II enzymes for which a number of polymorphisms have been described, and this effect may result in some interindividual variability in the concentrations of circulating metabolites (37). Polyphenol metabolites formed by the microbiome were also shown to be less suitable as biomarkers of intake than were other polyphenol metabolites because of the variability of the microbiome between individuals (38).

The annotation of biomarkers for each of the selected foods allowed for the attribution of the main clusters (metabolite-metabolite Pearson correlation coefficient >0.3) in the correlation heat map of the significant MS features to metabolites derived from the 6 foods (**Figure 3**), each cluster of which was explained by the co-occurrence of several characteristic polyphenols in each food. The largest cluster (986 MS features) gathered mass spectral features that were correlated to coffee intake and included all identified coffee metabolites (Table 1). The other clusters corresponded to red wine, citrus fruit, and tea and chocolate products with 656, 345, and 102 features, respectively. A small cluster of MS features derived from apples and pears was also recognized. A number of features were common to the red wine and tea clusters and included metabolites such as methylgallic acid (and its sulfate ester metabolites), which are known to be present in both dietary sources, or catechin and its gut microbial metabolites (Supplemental Tables 4–9) (36).

We also examined correlation coefficients between all MS features detected and habitual intakes of the 6 selected foods assessed through the use of dietary questionnaires at the same significance level of 1% with Benjamini-Hochberg adjusted $P$ values (35) and a minimum threshold value of 0.3 for the correlation value. Features correlated with habitual food intakes were identified for 5 of the 6 foods as follows: coffee: 817 features; red wine: 499 features; citrus fruit: 7 features; tea: 36 features; and apples and pears: 10 features. A large fraction of the features (70–91% for coffee, tea, red wine, and citrus fruit) that correlated with acute dietary intake were shown to also correlate with habitual intake as illustrated in the Venn diagrams (Figure 1B). However, the number of features related to habitual intake was significantly lower for all foods and particularly for tea, citrus fruit, and apples and pears.

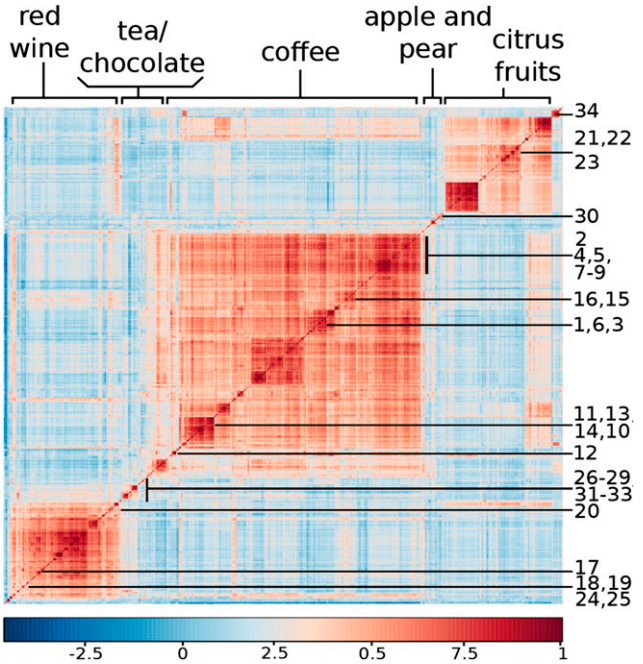


**FIGURE 3** Metabolite-metabolite correlation analysis of the 2272 mass spectral features correlated to dietary intakes of 6 polyphenol-rich foods (24-h dietary recalls). The heat map shows clusters for each food made of mass spectral features derived from metabolites co-occurring in a same food and excreted in the same urine samples. Numbers refer to the annotated polyphenol metabolites (Table 1).

### Polyphenol metabolome shows variable profiles according to country

O-PLS-DA models were used to compare metabolic profiles in the 4 countries by pairs. An examination of the loadings of the O-PLS-DA models showed that, of the features associated with the consumption of polyphenol-rich foods, many of the polyphenol markers were covariant with the country assessed. The most-contrasted O-PLS-DA model was obtained for Germany and Italy (**Figure 4**). Some of the dietary biomarkers annotated in this study, particularly those characteristic of coffee, tea, red wine, and citrus fruit intakes, significantly contribute to the loadings of the O-PLS-DA S-plots and to the discrimination of German and Italian populations (Figure 4). This result showed the relatively higher exposure of the German population to metabolites from coffee and tea (Figure 4A, D) and the higher exposure of the Italian population to red wine and citrus fruit metabolites (Figure 4B, C).

## DISCUSSION

The identification of associations between dietary exposure and disease outcomes relies on the accurate estimation of dietary intake. Dietary biomarkers have been increasingly used either to validate other dietary assessment tools such as questionnaires or to complement them (4, 39). Still, the number of biomarkers that have been identified and used in epidemiogic studies is relatively limited (3). In the current article, we showed that the application of metabolomics to a cross-sectional study with rich dietary information allowed for the identification of a large number of dietary compounds associated with food intake. This discovery approach necessitated rich databases on food constituents and their metabolites to annotate unknown signals correlated to the diet. Focus was put in this work on the polyphenol metabolome because of the availability of the Phenol-Explorer database that contains information on >500 dietary polyphenols and their food sources and on 350 polyphenol metabolites described in humans or experimental animals. More than 80 polyphenol metabolites associated with intakes of 6 polyphenol-rich foods were identified, some of which showed an excellent capacity to predict intakes of the selected foods (Figure 2). The chemical nature of 4 of these metabolites, together with the knowledge of their dietary precursors and the distribution in foods of these precursors (Supplemental Table 11), provided a high level of confidence about the biochemical parentage that links dietary exposure to the biomarker and contributes to their validation as biomarkers.

The current biomarker discovery approach required rich data on food constituents and their metabolites as shown in the Phenol-Explorer database and also necessitated high-quality data on dietary intake as was available in the EPIC calibration study (24). The relatively low number of biomarkers identified for apples and pears or for chocolate products could have equally reflected the insufficient accuracy of intake measurements for such foods used in a wide number of recipes made of varying amounts and the composition of ingredients or the lack of the specificity and sensitivity of the few phenolic biomarkers identified as previously discussed.

In this work, biomarkers that were correlated with both acute intakes, as measured with the use of 24-HDRs, and habitual intakes, as estimated with the use of a food-frequency questionnaire, were identified. Nutritional epidemiologists are particularly interested in biomarkers of habitual dietary exposure. However the number of MS features that correlated with habitual intake was significantly less than those that correlated with acute intake (Figure 2).

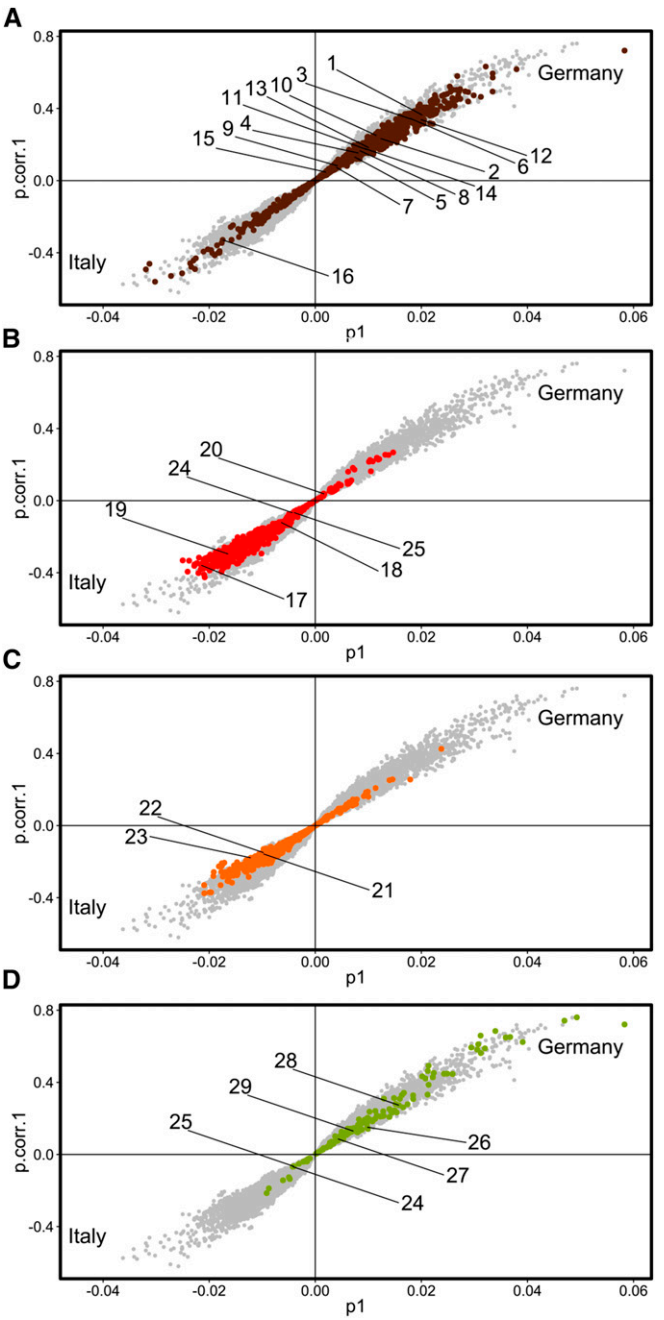


**FIGURE 4** Orthogonal partial least-squares discriminant analysis S plots showing the contribution of the mass spectral features measured in urine to the discrimination of the German ($n$ = 175) and Italian ($n$ = 178) populations. Biomarker characteristics of each selected polyphenol-rich food are highlighted in color. (A) Coffee. (B) Red wine. (C) Citrus fruit. (D) Tea. Numbers refer to the annotated polyphenol metabolites (Table 1). The model reveals the robust discrimination of the urinary mass spectral data from these 2 countries with contrasted typical dietary intakes [R2X: 0.21; R2Y: 0.94; Q2(cum): 0.89; 1 + 3 + 0 ($n$ predictive components + $n$ orthogonal components)]. R2X, the fraction of the variation of the X variables (i.e. MS features) explained by the model; R2Y, the fraction of the variation in the Y variable (i.e. country) explained by the model; Q2(cum): the cumulative predictive ability of the model, calculated as 1 − the predicted residual sum of squares (PRESS) divided by the sum of squares (SS).

A number of elements may explain this difference. First, habitual food-intake measurements may be less accurate than are acute intake measurements, mainly because of the longer period of dietary assessment (12 mo compared with 1 d) and the limited number

of food items included in the dietary questionnaires. Second, some foods are characterized by sporadic or seasonal consumption, thus leading to the quick elimination of polyphenol metabolites in urine, which is most often complete 24 h after ingestion (40). In agreement with this last hypothesis, the largest overlap between markers correlated to acute and habitual dietary intake measurements were observed for regularly consumed foods such as coffee, tea, and red wine (Figure 1B). Therefore, biomarkers identified in this work should be considered as biomarkers of acute food intake with applications in cohort studies largely limited to assess exposures of most regularly consumed foods.

The current metabolomic approach was also used to compare dietary exposure in individuals from the EPIC cohort who originated from 4 different European countries with great heterogeneity of dietary habits. Large differences in urinary metabolic profiles were observed between the 4 European countries. The most-contrasted O-PLS-DA model, when countries were compared by pairs, was obtained for Germany and Italy in agreement with the known contrast in dietary patterns in Northern and Southern Europe (Figure 4) (41). This contrast in the urinary MS data are likely largely explained by differences in diet as suggested by the contribution of the polyphenol metabolites to the O-PLS-DA models (Figure 4). We also checked that the biomarkers identified were not confounded by countries and truly reflected differences in dietary exposures. Indeed correlation coefficients between amounts of polyphenol metabolites and food intakes showed minor changes after adjustment for the country (**Supplemental Table 12**).

These differences in metabolic profiles between countries can be used to reveal country-specific differences in dietary exposure. A previous metabolomic study conducted in the International Study of Macro/micronutrients and Blood Pressure with the use of nuclear magnetic resonance spectroscopy revealed some differences in biomarker amounts that could be notably explained by differences in alcohol (ethanol and ethyl glycoside) and fish (trimethylamine-*N*-oxide) consumption in Japanese, Chinese, and American populations studied (42). However, many more dietary-related biomarkers can be detected with the far more–sensitive high-resolution MS technique used in the current work.

The value of a metabolomic approach to compare dietary exposures in different populations is further exemplified by coffee. A large number of phenolic metabolites identified in urine were shown to be characteristic of coffee consumption, and feruloylquinic acid, which is a major polyphenol characteristic of coffee (18, 43), was identified as one of the best predictors of coffee consumption (Table 1, assignment 3). Concentrations of feruloylquinic acid in urine showed country-specific differences, as shown in Figure 4A, for the German and Italian populations that were related to differences in coffee consumption. Self-reported coffee intake expressed as mL/d day was 5 times higher in German consumers (mean ± SD: 648 ± 378 mL/24 h; $n = 150$) than in Italian consumers (average: 134 ± 89 mL/24 h; $n = 161$), whereas mass spectral intensities of feruloylquinic acid were only 2-fold larger in German coffee consumers (peak-area counts: $2.37 \times 10^5 \pm 1.59 \times 10^5$) than in Italian coffee consumers (peak-area counts: $1.20 \times 10^5 \pm 8.66 \times 10^4$) coffee consumers. These differences in the ratio of coffee intake and biomarker concentrations between Germany and Italy likely reflected variations in the mode of preparation of coffee brews (mainly filter coffee in Germany and espresso in Italy) and the dilution level of coffee drinks. Biomarkers should provide a more objective and reliable estimate of the exposure to food constituents and improve the subject classification in epidemiologic studies that aim to unravel the effects of coffee on disease risk.

Strengths of our study were the diversity of diets in the 4 populations studied, the availability of 24-h urine samples, the wealth and quality of the dietary information collected, the high sensitivity of the analytic method used to measure the food metabolome, and the use of the Phenol-Explorer database to annotate the food metabolome. Most often in cohort studies, only spot urine samples are available when urine samples have been collected. The combination of 24-h urine samples with very-rich dietary information collected on a same day made this study quite unique. The dataset generated in the current work can be mined to identify biomarkers for all sorts of foods and other lifestyle exposures. Our study also had some limitations including missing details on intakes for some foods (e.g., chocolate products) and the tentative identification of some biomarkers that will need to be confirmed when authentic standards are available. Another important limitation was that the design of the current study did not allow us to make the difference between acute biomarkers rapidly eliminated in the urine or bile and habitual biomarkers with longer half-lives in the organism. Habitual biomarkers may still be identified through the comparison of metabolic profiles with habitual dietary intake data, but the risk of confounding with other associated foods should carefully be considered.

In conclusion, the current work shows the major contribution of the food metabolome to the human metabolome, more particularly in urine as a principal route of excretion for by-products of digestion. Each ingested food contributes with its own metabolome made of thousands of nutrients and other chemicals, and this makes the food metabolome one of the most complex fractions of the human metabolome (3, 13). Measurement of the food metabolome in future epidemiologic studies should complement or substitute traditional methods based on questionnaires to improve dietary exposure assessment. Finally, measurement of the food metabolome should facilitate the provision of clearer evidence on the relations between dietary exposure, food composition, and risk of major chronic diseases such as cancer, cardiovascular diseases, or diabetes and shed new light on the causes of such diseases.

The authors' responsibilities were as follows—WMBE, PF, JAR, and AS: data interpretation; WMBE and DKB: chemical data collection and data preprocessing; WMBE and PF: statistical analyses; WMBE and AS: writing of the manuscript; AS, SR, NS, MJ, and IR: study concept and design; AS: primary responsibility for the final content of the manuscript; PF, SR, NS, CB, MJ, FC-C, GF, M-CB-R, VAK, TK, HB, AT, PL, DT, DP, SG, RT, PV, AM, and IR: recruitment, dietary data collection, biological sample collection, and follow-up or management of the EPIC cohort; and all authors: critical revision and approval of the final version of the manuscript. None of the authors reported a conflict of interest related to the study.

## REFERENCES

1. 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:D801–7.
2. University of Alberta. [Internet]. FooDB (cited 2013 Sep 3). Available from: http://www.foodb.ca/.
3. Scalbert A, Brennan L, Manach C, Andres-Lacueva C, Dragsted LO, Draper J, Rappaport SM, van der Hooft JJJ, Wishart DS. The food metabolome – a window over dietary exposure. Am J Clin Nutr 2014; 99:1286–308.
4. 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.

5. Angerer J, Ewers U, Wilhelm M. Human biomonitoring: state of the art. Int J Hyg Environ Health 2007;210:201–28.
6. 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 2014;100:11–26.
7. 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.
8. 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.
9. Rothwell JA, Fillâtre Y, Martin J-F, Lyan B, Pujos-Guillot E, Fezeu L, Hercberg S, Comte B, Galan P, Touvier M, et al. New biomarkers of coffee consumption identified by the non-targeted metabolomic profiling of cohort study subjects. PLoS One 2014;9:e93474.
10. 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.
11. 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.
12. 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.
13. Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, Bjorndahl TC, Krishnamurthy R, Saleem F, Liu P, et al. The human urine metabolome. PLoS One 2013;8:e73076.
14. Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. Stat Med 1989;8:1051–69, discussion 71–3.
15. Ferrari P, Roddam A, Fahey MT, Jenab M, Bamia C, Ocke M, Amiano P, Hjartaker A, Biessy C, Rinaldi S, et al. A bivariate measurement error model for nitrogen and potassium intakes to evaluate the performance of regression calibration in the European Prospective Investigation into Cancer and Nutrition study. Eur J Clin Nutr 2009;63 (Suppl 4):S179–87.
16. Kipnis V, Subar AF, Midthune D, Freedman LS, Ballard-Barbash R, Troiano RP, Bingham S, Schoeller DA, Schatzkin A, Carroll RJ. Structure of dietary measurement error: results of the OPEN biomarker study. Am J Epidemiol 2003;158:14–21.
17. 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.
18. Pérez-Jiménez J, Fezeu L, Touvier M, Arnault N, Manach C, Hercberg S, Galan P, Scalbert A. Dietary intake of 337 polyphenols in French adults. Am J Clin Nutr 2011;93:1220–8.
19. Pérez-Jiménez J, Neveu V, Vos F, Scalbert A. Systematic analysis of the content of 502 polyphenols in 452 foods and beverages: an application of the Phenol-Explorer database. J Agric Food Chem 2010;58: 4959–69.
20. 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.
21. 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.
22. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, Charrondiere UR, Hemon B, Casagrande C, Vignat J, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr 2002;5: 1113–24.
23. Slimani N, Bingham S, Runswick S, Ferrari P, Day NE, Welch AA, Key TJ, Miller AB, Boeing H, Sieri S, et al. Group level validation of protein intakes estimated by 24-hour diet recall and dietary questionnaires against 24-hour urinary nitrogen in the European Prospective Investigation into Cancer and Nutrition (EPIC) calibration study. Cancer Epidemiol Biomarkers Prev 2003;12:784–95.
24. Slimani N, Ferrari P, Ocke M, Welch A, Boeing H, van Liere M, Pala V, Amiano P, Lagiou A, Mattisson I, et al. Standardization of the 24-hour diet recall calibration method used in the European Prospective Investigation into Cancer and Nutrition (EPIC): general concepts and preliminary results. Eur J Clin Nutr 2000;54:900–17.
25. Jacobs S, Kroger J, Floegel A, Boeing H, Drogan D, Pischon T, Fritsche A, Prehn C, Adamski J, Isermann B, et al. Evaluation of various biomarkers as potential mediators of the association between coffee consumption and incident type 2 diabetes in the EPIC-Potsdam Study. Am J Clin Nutr 2014;100:891–900.
26. Zamora-Ros R, Luján-Barroso L, Bueno-de-Mesquita HB, Dik VK, Boeing H, Steffen A, Tjønneland A, Olsen A, Bech BH, Overvad K, et al. Tea and coffee consumption and risk of esophageal cancer: the European Prospective Investigation into Cancer and Nutrition study. Int J Cancer 2014;135:1470–9.
27. Hooper L, Kay C, Abdelhamid A, Kroon PA, Cohn JS, Rimm EB, Cassidy A. Effects of chocolate, cocoa, and flavan-3-ols on cardiovascular health: a systematic review and meta-analysis of randomized trials. Am J Clin Nutr 2012;95:740–51.
28. 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.
29. Chiva-Blanch G, Arranz S, Lamuela-Raventos RM, Estruch R. Effects of wine, alcohol and polyphenols on cardiovascular disease risk factors: evidences from human studies. Alcohol Alcohol 2013;48:270–7.
30. Edmands WM, Ferrari P, Scalbert A. Normalization to specific gravity prior to analysis improves information recovery from high resolution mass spectrometry metabolomic profiles of human urine. Anal Chem 2014;86:10925–31.
31. 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.
32. Edmands WM, Barupal DK, Scalbert A. MetMSLine: an automated and fully integrated pipeline for rapid processing of high-resolution LC-MS metabolomic datasets. Bioinformatics 2015;31:788–90.
33. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Ann Stat 2001;29:1165–88.
34. 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.
35. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. J Roy Stat Soc B Met 1995;57:289–300.
36. Neveu V, Perez-Jiménez J, Vos F, Crespy V, du Chaffaut L, Mennen L, Knox C, Eisner R, Cruz J, Wishart D, et al. Phenol-Explorer: an online comprehensive database on polyphenol contents in foods. Database (Oxford) 2010;2010:bap024.
37. Manach C, Scalbert A, Morand C, Rémésy C, Jimenez L. Polyphenols - food sources and bioavailability. Am J Clin Nutr 2004;79:727–47.
38. 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.
39. Kuhnle GGC. Nutritional biomarkers for objective dietary assessment. J Sci Food Agric 2012;92:1145–9.
40. 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.
41. Slimani N, Fahey M, Welch AA, Wirfalt E, Stripp C, Bergstrom E, Linseisen J, Schulze MB, Bamia C, Chloptsios Y, et al. Diversity of dietary patterns observed in the European Prospective Investigation into Cancer and Nutrition (EPIC) project. Public Health Nutr 2002;5:1311–28.
42. Dumas ME, Maibaum EC, Teague C, Ueshima H, Zhou BF, Lindon JC, Nicholson JK, Stamler J, Elliott P, Chan Q, et al. Assessment of analytical reproducibility of H-1 NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP study. Anal Chem 2006;78:2199–208.
43. 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.


The American Journal of Clinical Nutrition