Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cancer[1–5]

Kristin A Guertin, Erikka Loftfield, Simina M Boca, Joshua N Sampson, Steven C Moore, Qian Xiao, Wen-Yi Huang, Xiaoqin Xiong, Neal D Freedman, Amanda J Cross, and Rashmi Sinha

ABSTRACT

Background: Coffee intake may be inversely associated with colorectal cancer; however, previous studies have been inconsistent. Serum coffee metabolites are integrated exposure measures that may clarify associations with cancer and elucidate underlying mechanisms.

Objectives: Our aims were 2-fold as follows: *1*) to identify serum metabolites associated with coffee intake and *2*) to examine these metabolites in relation to colorectal cancer.

Design: In a nested case-control study of 251 colorectal cancer cases and 247 matched control subjects from the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial, we conducted untargeted metabolomics analyses of baseline serum by using ultrahigh-performance liquid-phase chromatography–tandem mass spectrometry and gas chromatography–mass spectrometry. Usual coffee intake was self-reported in a food-frequency questionnaire. We used partial Pearson correlations and linear regression to identify serum metabolites associated with coffee intake and conditional logistic regression to evaluate associations between coffee metabolites and colorectal cancer.

Results: After Bonferroni correction for multiple comparisons ($P = 0.05 \div 657$ metabolites), 29 serum metabolites were positively correlated with coffee intake (partial correlation coefficients: 0.18–0.61; $P < 7.61 \times 10^{-5}$); serum metabolites most highly correlated with coffee intake (partial correlation coefficients >0.40) included trigonelline (*N′*-methylnicotinate), quinate, and 7 unknown metabolites. Of 29 serum metabolites, 8 metabolites were directly related to caffeine metabolism, and 3 of these metabolites, theophylline (OR for 90th compared with 10th percentiles: 0.44; 95% CI: 0.25, 0.79; *P*-linear trend = 0.006), caffeine (OR for 90th compared with 10th percentiles: 0.56; 95% CI: 0.35, 0.89; *P*-linear trend = 0.015), and paraxanthine (OR for 90th compared with 10th percentiles: 0.58; 95% CI: 0.36, 0.94; *P*-linear trend = 0.027), were inversely associated with colorectal cancer.

Conclusions: Serum metabolites can distinguish coffee drinkers from nondrinkers; some caffeine-related metabolites were inversely associated with colorectal cancer and should be studied further to clarify the role of coffee in the cause of colorectal cancer. The Prostate, Lung, Colorectal, and Ovarian trial was registered at clinicaltrials.gov as NCT00002540. *Am J Clin Nutr* 2015;101:1000–11.

Keywords: coffee, colorectal cancer, dietary intake, metabolomics, metabolites

INTRODUCTION

Coffee is widely consumed globally and is a rich source of bioactive compounds that may favorably affect cancer risk; these compounds include polyphenols (e.g., chlorogenic acids), diterpenes (e.g., cafestol), and caffeine (1–3). Coffee intake has been inversely associated with all-cause mortality (4–6), cause-specific mortality because of heart disease, respiratory disease, stroke, and diabetes (7), Parkinson's disease (8), and cancers of the endometrium (9, 10), liver (11, 12), skin (13, 14), and colon (15).

Colorectal cancer is the fourth most commonly diagnosed cancer and the second leading cause of cancer mortality in the United States (16). A recent study from the large NIH-AARP Diet and Health cohort reported lower risk of colon cancer, particularly of proximal tumors, in coffee consumers (15). Moreover, previous meta-analyses of case-control and cohort studies supported an inverse association between coffee drinking and colorectal cancer (17–19). However, a pooled analysis of 13 prospective studies reported no association between coffee intake and colon cancer (20). The measurement error associated with self-report as well as differences in coffee composition that are due to the bean type, roast, and preparation method may in

[1] From the Nutritional Epidemiology Branch (KAG, EL, SCM, QX, NDF, and RS), the Biostatistics Branch (JNS), and the Occupational and Environmental Epidemiology Branch (W-YH), Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Department of Health and Human Services, Bethesda, MD; the Innovation Center for Biomedical Informatics and Department of Oncology, Georgetown University Medical Center, Washington, DC (SMB); Information Management Services Inc., Silver Spring, MD (XX); and the Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine, Imperial College London, St. Mary's Campus, Norfolk Place, London, United Kingdom (AJC).

[2] KAG and EL contributed equally to this article as co-first authors.

[3] Supplemental Tables 1 and 2 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.

[4] Supported by the Intramural Research Program of the National Cancer Institute, NIH.

[5] Address correspondence to KA Guertin and E Loftfield, National Cancer Institute, NIH, 9609 Medical Center Drive, Room 6E326, Mail Stop Code 9760, Bethesda, MD 20892. E-mail: kristin.guertin@nih.gov and erikka.loftfield@nih.gov.

Received July 22, 2014. Accepted for publication January 26, 2015.

First published online March 11, 2015; doi: 10.3945/ajcn.114.096099.

part account for these inconsistent findings. In addition, individual differences in the metabolism of coffee that are due to genetic (21, 22) or microbiomic (23) variation may affect the bioavailability of key metabolites that play a role in the cause of disease. Metabolomics, which is the study of metabolite dynamics, is an innovative strategy for studying diet and human health because it potentially measures an intermediate phenotype that integrates intake, genotype, gut microbial metabolism, and other host factors and may elucidate underlying biological mechanisms of disease (24).

Previously, a case-control study nested within the Prostate, Lung, Colorectal, and Ovarian (PLCO)[6] Cancer Screening Trial (clinicaltrials.gov; NCT00002540) was used to explore the cross-sectional associations of serum metabolites with usual dietary intake (25) and, later, to agnostically investigate metabolites prospectively associated with risk of colorectal cancer (26). The current analysis was based on the same nested case-control study, but it was hypothesis-driven and focused on metabolites specifically associated with self-reported habitual coffee intake. Metabolites associated with coffee were examined in relation to colorectal cancer. Furthermore, this analysis expands on the scope of the previously reported diet and metabolite analysis by including metabolites of unknown identity.

METHODS

PLCO Cancer Screening Trial

The PLCO Cancer Screening Trial is a multicenter randomized screening trial of prostate, lung, colorectal, and ovarian cancers (27, 28). More than 155,000 participants in the United States aged 55–74 y with no history of the 4 named cancers were randomly assigned from 1993 to 2001 to a screening or control arm. The current investigation was restricted to individuals in the screening arm of the trial (n = 77,445).

At baseline, participants in the screening arm were offered a flexible sigmoidoscopy to screen for colorectal cancer in the distal colorectum; 83% of participants (n = 64,658) in the screening arm underwent the procedure, and 89% of these procedures (n = 57,559) were considered successful (insertion to ≥50 cm with >90% of mucosa visible or a suspect lesion identified). Participants with identified neoplastic lesions were referred to their health care provider for a colonoscopy. All participants in the screening arm underwent a follow-up flexible sigmoidoscopy either 3 or 5 y after baseline. The study was approved by the Institutional Review Board at the National Cancer Institute and the 10 study centers; all participants provided written informed consent.

Study sample

Eligible participants completed baseline dietary and risk factor questionnaires, consented for biospecimen use, and were free of colorectal cancer at baseline (n = 52,705). We excluded participants with a self-reported personal history of cancer (except basal cell skin cancer) (n = 4924), <6 mo of follow-up (an additional 168 individuals), a rare cancer during follow-up (an additional 1074 individuals), self-reported colon disease (Crohn disease, ulcerative colitis, familial polyposis, Gardner syndrome, or colorectal polyps; an additional 6429 individuals), and participants who did not have a baseline serum sample available (an additional 2866 individuals); some participants met more than one of the exclusion criteria.

Case ascertainment

Incident colorectal cancer cases were identified through February 2011 by self-reported annual health questionnaires and linkage to the National Death Index and histologically confirmed through a medical record review. Cases included individuals with a first primary incident colorectal adenocarcinoma [International Classification of Diseases for Oncology (29), ICD-O-3 codes: C180–189, C199, C209, and C260, excluding morphologies 8240–8249] identified ≥6 mo after baseline. Control subjects (n = 254) were free from any cancer at the time matched cases were diagnosed and incidence-density matched to cases for age (5-y intervals), year of randomization, season of blood draw, sex, and race. For this analysis, we further excluded cases or controls with missing information on baseline coffee intake (n = 7) or BMI (in kg/m^2; n = 4). Our final analytic sample (n = 498) included 251 cases and 247 controls.

Exposure assessment

At baseline, participants completed a risk-factor questionnaire and 137-item food-frequency questionnaire (FFQ) that was based on 2 previously validated FFQs (30, 31) and queried about the usual portion size and frequency of food and beverage consumption, including total coffee intake, over the past year; the baseline FFQ did not differentiate between caffeinated and decaffeinated coffee. In coffee drinkers, the frequency of intake (<1 time/mo to ≥6 times/d) was multiplied by a gram amount, which was dependent on the sex of the subject and response to the serving size (small, medium, or large cup); gram amounts came from the USDA's 1994–1996 Continuing Survey of Food Intakes by Individuals database (32). Grams of coffee were converted to the number of medium (i.e., 12 oz) cups per day.

Type of coffee intake, caffeinated or decaffeinated, was measured by the National Cancer Institute Diet History Questionnaire (DHQ), which was administered in December 1998 ~3 y after baseline (range: 2–9 y) (33). Consequently, all analyses, except the secondary analysis that considered coffee type, defined coffee intake as measured at baseline by using an FFQ.

Metabolite assessment

Baseline serum samples were nonfasting, and the time of blood draw ranged from 0700 to 1600, with ~45% and ~82% of blood draws occurring before 1000 and 1300, respectively. An untargeted metabolomic profiling analysis was performed on a baseline serum sample from each individual by Metabolon Inc. as previously described (34–36). In brief, serum samples, which had been stored at −70°C and not previously thawed, were subjected to an untargeted single methanol extraction followed by protein precipitation (37). Metabolites were detected by using ultrahigh-performance liquid-phase chromatography–tandem mass spectrometry and gas chromatography–mass spectrometry and identified by comparison to

[6] Abbreviations used: DHQ, Diet History Questionnaire; FFQ, food-frequency questionnaire; LLOD, lower limit of detection; PLCO, Prostate, Lung, Colorectal, and Ovarian.

chemical reference libraries. Mass spectral peaks, retention times, and *m/z* were used to determine the relative quantities of each metabolite.

Each batch was run in a single day and contained ≤30 samples including blinded quality-control samples of pooled serum at a level of 10%. Matched cases and controls were arranged consecutively within a batch; the order of cases compared with controls was counterbalanced within each batch. Every sixth sample, Metabolon Inc. inserted a standard for quality-control purposes.

We normalized metabolites within each batch by dividing an individual's metabolite value by the batch mean of all nonmissing values for a given metabolite; batch-normalized metabolites were ln transformed [i.e., ln(metabolite)]. Metabolite values below the lower limit of detection (LLOD) were assigned the minimum of all observed values.

Statistical analyses

Our first objective was to measure associations between self-reported coffee consumption and metabolites. We considered baseline coffee consumption as both a categorical variable and continuous variable. For the categorical variable, we divided individuals into groups of no, low, and high coffee intakes by using the median intake in consumers to differentiate low and high groups. For the continuous variable, we considered ln(coffee intake + c), where c is a constant defined to be one-third of the minimum reported intake.

We modeled associations between coffee and metabolites by using linear regression models adjusted for sex, tobacco smoking status (current, former, or never smokers of any cigarettes, pipes, or cigars), age (continuous), and current BMI (continuous). The threshold for statistical significance was set on the basis of Bonferroni correction for the number of detected metabolites at $P = 0.05$ ($0.05 \div 657$ metabolites); metabolites with linear regression P values below this limit were considered candidate coffee biomarkers. We examined partial Pearson correlations between ln(coffee) and ln(metabolite) for each candidate coffee biomarker with adjustments for sex, smoking, age, and BMI. To allow for the examination of trends in relative metabolite levels across increasing categories of coffee intake, we rank ordered the 498 subjects by metabolite level for each candidate coffee biomarker and calculated percentiles by dividing each ranking by the total number of values (e.g., the 249th of 498th ranked metabolite values corresponded to the 50th percentile). We calculated the mean percentile of a given metabolite distribution within each coffee intake category. We also created a metabolite heat map with the R program (version 2.15.2) using hierarchical clustering (R Foundation for Statistical Computing) to illustrate correlations between metabolites that were statistically significantly associated with coffee.

Because smoking is strongly correlated with coffee drinking and may affect coffee metabolism, we examined the partial Pearson correlations between ln(metabolite) and ln(coffee), stratified by smoking status (current, former, or never). In addition, the time of day that nonfasting blood samples were collected may affect correlations between self-reported coffee intake and metabolites; therefore, we examined partial Pearson correlations between ln (metabolite) and ln(coffee) stratified by the time of blood draw (0700 until 1000, 1000 until 1300, or 1300 until 1600).

In secondary analyses, we examined partial Pearson correlations between ln(metabolite) and ln(coffee) for caffeinated coffee only, decaffeinated coffee only, and mixed (a combination of caffeinated and decaffeinated) as measured by the DHQ administered during follow-up. An additional 20 individuals who were missing data on coffee intake on the DHQ were excluded from these analyses, which left 478 individuals. The 3 coffee-consumption categories (caffeinated only, decaffeinated only, and mixed) were mutually exclusive; subjects who consumed no coffee ($n = 72$) were included in all 3 analyses.

Our second objective was to evaluate associations between the candidate coffee biomarkers, identified herein, and colorectal cancer. We used conditional logistic regression models adjusted for smoking status, age, and BMI to estimate ORs and 95% CIs for coffee-associated metabolites and colorectal cancer. Additional adjustments for alcohol intake, nutrient-density adjusted red and white meats, dietary fat, and total energy intake (kcal/d) did not appreciably alter regression coefficients in that no regression coefficient changed >3%. For metabolites for which ≥20% of the sample had undetectable levels, we compared individuals with low (i.e., detectable but equal to or below the median) and high (i.e., above the median) relative metabolite levels to those with undetectable levels. To have an adequately sized reference group, for metabolites for which <20% of the sample had undetectable levels for a given metabolite, we compared individuals with high levels of the metabolites with those with low or undetectable levels. P-linear trend values were calculated by using the continuous variable ln(metabolite), and continuous estimates were reported as ORs for 90th compared with 10th percentiles. Because of our a priori hypotheses that coffee-associated serum metabolites are inversely associated with colorectal cancer, $P < 0.05$ was considered significant.

We assessed the technical reliability of our data by using intraclass correlation coefficients for quality-control samples. In addition, the overall reliability and validity of this platform was previously reported (38). Unless otherwise specified, all analyses were conducted with SAS 9.1.3 software (SAS Institute).

RESULTS

Baseline characteristics of cases and controls are presented in **Table 1**. A wide range of coffee intake was reported, with >10% of individuals drinking ≥6 cups/d. High coffee drinkers (≥2.5 cups/d) had median intake of ~4.3 cups/d, and the majority of these individuals (79%) were men. Current and former smoking was more prevalent in individuals who consumed larger amounts of coffee, and median alcohol intake was also higher in high coffee drinkers.

We identified a total of 657 unique metabolites in human serum; of these, 428 metabolites were known, whereas 229 metabolites were of unknown identity. After Bonferroni adjustment for multiple testing ($P = 0.05 \div 657$ metabolites), 29 serum metabolites were significantly associated with self-reported coffee intake at $P < 7.61 \times 10^{-5}$ (**Table 2**); all associations were positive such that individuals with higher coffee intake had higher relative metabolite levels. Metabolites most strongly associated with self-reported coffee drinking, all with partial correlations >0.40, were trigonelline (*N′*-methylnicotinate), quinate, and 7 metabolites of unknown identity. Eight additional serum metabolites [paraxanthine, *N*-(2-furoyl) glycine, catechol sulfate, caffeine, 1-methylxanthine,

TABLE 1
Demographic characteristics of participants ($n = 498$) in a metabolomics study nested within the PLCO Cancer Screening Trial, stratified by self-reported total coffee intake[1]

	Coffee intake			
	None	Low (<2.5 cups/d)	High (≥2.5 cups/d)	*P*
n	54	236	208	
Age, y	65.1 ± 5.0[2]	68.0 ± 5.3	68.2 ± 5.1	<0.001
Caucasian, *n* (%)	50 (93)	203 (86)	191 (92)	0.040
Women, *n* (%)	33 (61)	142 (60)	44 (21)	<0.0001
Education,[3] *n* (%)				0.371
High school or less	21 (39)	72 (31)	74 (36)	
Post–high school/some college	19 (35)	74 (31)	63 (30)	
College/postgraduate	13 (24)	90 (38)	71 (34)	
Smoking status,[4] *n* (%)				<0.001
Current	3 (6)	12 (5)	37 (18)	
Former	15 (28)	114 (48)	118 (57)	
Never	36 (67)	110 (47)	55 (26)	
BMI, kg/m^2	28.3 ± 5.6	26.7 ± 4.5	27.7 ± 4.4	0.019
Alcohol intake, g/d	0 (0–79)[5]	1 (0–322)	4 (0–173)	0.112
Coffee intake				<0.001
g/d	0	337 (2–843)	1546 (875–6295)	
Cups/d	0	1.0 (<0.1 to 2.4)	6.5 (2.5–17.7)	

[1]Coffee intake was assessed by using a food-frequency questionnaire at baseline and measured in grams of total coffee per day (includes caffeinated and decaffeinated); percentages may not add to 100 because of missing data or rounding. Differences between coffee-intake groups were assessed by using ANOVA and chi-square tests for continuous and categorical variables, respectively. PLCO, Prostate, Lung, Colorectal, and Ovarian.

[2]Mean ± SD (all such values).

[3]One participant had missing information on education.

[4]Self-reported smoking of any tobacco (cigarettes, pipes, or cigars).

[5]Median; range in parentheses (all such values).

theophylline, and 2 unknown metabolites] were moderately correlated (partial correlations: 0.30–0.38) with self-reported coffee intake. Additional metabolites related to the degradation of caffeine (e.g., 1,3-dimethylurate and 7-methylxanthine), chlorogenic acids (e.g., hippurate), and trigonelline (e.g., nicotinamide) that did not withstand Bonferroni correction did meet nominal significance ($P < 0.05$) (data not shown).

For 29 metabolites significantly associated with self-reported coffee intake, the percentage of participants who had levels below the LLOD ranged from 1% to 86% (Table 2). The mean metabolite-level percentile was lowest in those who reported no coffee consumption and increased with increasing coffee intake; e.g., non-, low, and high coffee drinkers had mean trigonelline (*N′*-methylnicotinate) levels in the 17th, 44th, and 65th percentiles of the distribution, respectively. Similarly, the percentage of participants with metabolite levels below the LLOD was consistently lowest in high consumers and highest in nonconsumers of coffee (**Supplemental Table 1**).

We also evaluated correlations between each of the 29 metabolites significantly associated with self-reported coffee intake, which are shown as a heat map in **Figure 1**. All of the top coffee metabolites were positively correlated with each other; correlations between caffeine, theophylline, paraxanthine, and 1,7-dimethylurate were especially strong (>0.7). In a sensitivity analysis, we calculated the principal components on the basis of the 29 significant metabolites. In a linear model that already included age, sex, smoking status, and BMI, the addition of the first principal component explained an additional 13% of the variation in coffee consumption, which was less than that explained by adding trigonelline only (34%). With the use of the adjusted R^2 as our metric for model fit, the addition of further principal components did not improve the model fit.

Correlations between total coffee intake and metabolites were similar when stratified by smoking status (**Supplemental Table 2**). With few current smokers ($n = 52$), only the metabolite trigonelline withstood Bonferroni correction for multiple testing; however, all 29 metabolites were nominally significantly associated with coffee intake in these stratified analyses. No consistent pattern was observed when correlations between total coffee intake and metabolites were stratified by time of blood draw (data not shown).

In secondary analyses, we evaluated correlations of caffeinated coffee only (185 individuals), decaffeinated coffee only (110 individuals), and mixed coffee consumption (111 individuals) with serum metabolites (**Table 3**); nondrinkers (72 individuals) were included in each of the 3 stratified analyses. Each of the 29 metabolites identified in Table 2 was also significantly associated ($P < 7.61 \times 10^{-5}$) with caffeinated-only coffee intake, with the exception of theobromine, which neared significance ($P = 9.13 \times 10^{-5}$). Mixed coffee and decaffeinated coffee intakes were significantly associated with 22 and 13, respectively, of the 29 metabolites associated with total coffee; significantly associated metabolites for mixed and decaffeinated coffee included trigonelline and quinate, which are not caffeine metabolites. Decaffeinated coffee was not associated with caffeine-related metabolites (e.g., paraxanthine), and metabolite correlations were weaker for decaffeinated coffee than for caffeinated only or mixed coffee intake. A narrower range of intake in the amount of coffee consumed by decaffeinated compared with caffeinated coffee drinkers may have

TABLE 2
Serum metabolites associated with self-reported total coffee intake in a nested case-control study within the PLCO Cancer Screening Trial (n = 498 participants)[1]

Serum metabolite	Mean percentile of metabolite distribution by level of coffee intake[2]			Participants with metabolites below the LLOD, n (%)	Partial correlation coefficient	P
	None (n = 54)	Low (n = 236)	High (n = 208)			
Cases/controls, n	26/28	118/118	107/101	—	—	—
Trigonelline (N'-methylnicotinate)	17.0	44.2	65.0	111 (22)	0.608	4.26×10^{-51}
Quinate	17.7	44.2	64.7	95 (19)	0.585	1.62×10^{-46}
X_12039	22.7	44.2	63.4	175 (35)	0.507	1.26×10^{-33}
X_13741	26.5	43.1	63.7	169 (34)	0.455	1.42×10^{-26}
X_12816	23.8	43.0	64.5	231 (46)	0.452	3.47×10^{-26}
X_14465	22.1	44.0	63.9	99 (20)	0.451	4.87×10^{-26}
X_14473	22.1	44.3	63.9	99 (20)	0.451	4.87×10^{-26}
X_17185	27.0	43.3	63.4	199 (40)	0.429	1.84×10^{-23}
X_12230	25.3	44.3	62.6	195 (39)	0.423	8.61×10^{-23}
Paraxanthine	30.8	42.7	63.0	29 (6)	0.383	1.12×10^{-18}
N-(2-furoyl)glycine	27.0	46.1	60.2	209 (42)	0.355	4.22×10^{-16}
Catechol sulfate	28.3	44.7	61.4	3 (1)	0.352	7.51×10^{-16}
Caffeine	32.6	45.1	59.8	19 (4)	0.327	1.02×10^{-13}
X_05426	29.3	44.2	61.7	21 (4)	0.324	1.6×10^{-13}
1-Methylxanthine	33.8	44.2	60.6	194 (39)	0.315	7.7×10^{-13}
Theophylline	32.9	44.4	60.6	61 (12)	0.306	4.06×10^{-12}
X_12329	34.6	45.8	58.5	321 (65)	0.302	7.22×10^{-12}
1,3,7-Trimethylurate	37.2	45.1	58.7	284 (57)	0.266	2.03×10^{-9}
3-Hydroxyhippurate	35.3	48.9	55.0	292 (59)	0.221	7.04×10^{-7}
1,7-Dimethylurate	40.3	44.9	58.1	227 (46)	0.219	9.52×10^{-7}
1-Methylurate	40.7	43.8	59.4	264 (53)	0.210	2.7×10^{-6}
X_14291	32.6	46.4	58.3	99 (20)	0.202	6.13×10^{-6}
X_14374	36.6	46.3	57.4	15 (3)	0.202	6.5×10^{-6}
Cyclo(leu-pro)	35.0	48.4	55.5	257 (52)	0.199	8.14×10^{-6}
4-Vinylphenol sulfate	34.5	46.1	58.2	34 (7)	0.188	2.65×10^{-5}
3-(3-Hydroxyphenyl) propionate	43.2	49.7	52.1	430 (86)	0.183	4.27×10^{-5}
Theobromine	46.1	45.8	55.6	20 (4)	0.182	4.61×10^{-5}
X_12734	37.4	46.1	57.6	254 (51)	0.179	6.16×10^{-5}
Cinnamoylglycine	41.9	50.4	51.4	335 (67)	0.178	7.37×10^{-5}

[1]Only significant associations from the linear regression of ln(metabolite) on ln(coffee) are shown; self-reported coffee intake was reported at baseline on the baseline food-frequency questionnaire; metabolites with the prefix X_ are of unknown identity. Partial correlations were determined between ln(metabolite) and ln(coffee) adjusted for sex, tobacco smoking status (current, former, and never), age (continuous), and BMI (continuous). P values are from the linear regression of ln(metabolite) on ln(coffee) adjusted for sex, tobacco smoking status (current, former, and never), age (continuous), and BMI (continuous). Bonferroni-corrected level of significance, $P < 7.61 \times 10^{-5}$. LLOD, lower limit of detection; PLCO, Prostate, Lung, Colorectal, and Ovarian.

[2]Average percentile of overall metabolite distribution represented by each category of coffee consumption; the average percentile was calculated by assigning a ranking for each participant for each metabolite of interest and dividing by the total number of individuals to convert the ranking into a percentile; within each category of coffee intake, percentiles were averaged for a given metabolite to calculate an average percentile. None, low, and high coffee intakes were defined as 0, <2.5, and ≥2.5 cups/d, respectively.

partially accounted for the weaker correlations between self-reported decaffeinated coffee intake and metabolites; median intakes (ranges) reported were 1.1 cups/d (<0.1–7.5 cups/d) and 3.6 cups/d (<0.1–10.1 cups/d), respectively.

Next, we investigated associations between the 29 coffee-associated serum metabolites identified herein and colorectal cancer. For metabolites for which <20% of the sample had undetectable levels for a given metabolite, we compared individuals with high levels of the metabolite with those with low or undetectable levels (**Table 4**). Individuals with higher compared with lower levels of theophylline (OR for 90th compared with 10th percentiles: 0.44; 95% CI: 0.25, 0.79; P-linear trend = 0.006), caffeine (OR for 90th compared with 10th percentiles: 0.56; 95% CI: 0.35, 0.89; P-linear trend = 0.015), and paraxanthine (OR for 90th compared with 10th percentiles: 0.58; 95% CI: 0.36, 0.94; P-linear trend = 0.027) had lower risk of colorectal cancer. For metabolites for which ≥20% of the sample had undetectable levels, we compared individuals with low (i.e., detectable levels but below the median) and high (i.e., above the median) metabolite levels to those with undetectable levels (**Table 5**); no significant associations were observed in this set of metabolites. Similarly, the modest inverse association between self-reported coffee intake (OR for high compared with no consumption: 0.85; 95% CI: 0.43, 1.72; OR for low compared with no consumption: 0.90; 95% CI: 0.46, 1.74; P-linear trend = 0.453) was NS.

DISCUSSION

We identified 29 serum metabolites associated with self-reported usual coffee intake in free-living US adults. Of these metabolites, theophylline, caffeine, and paraxanthine were inversely associated with colorectal cancer. In contrast and as

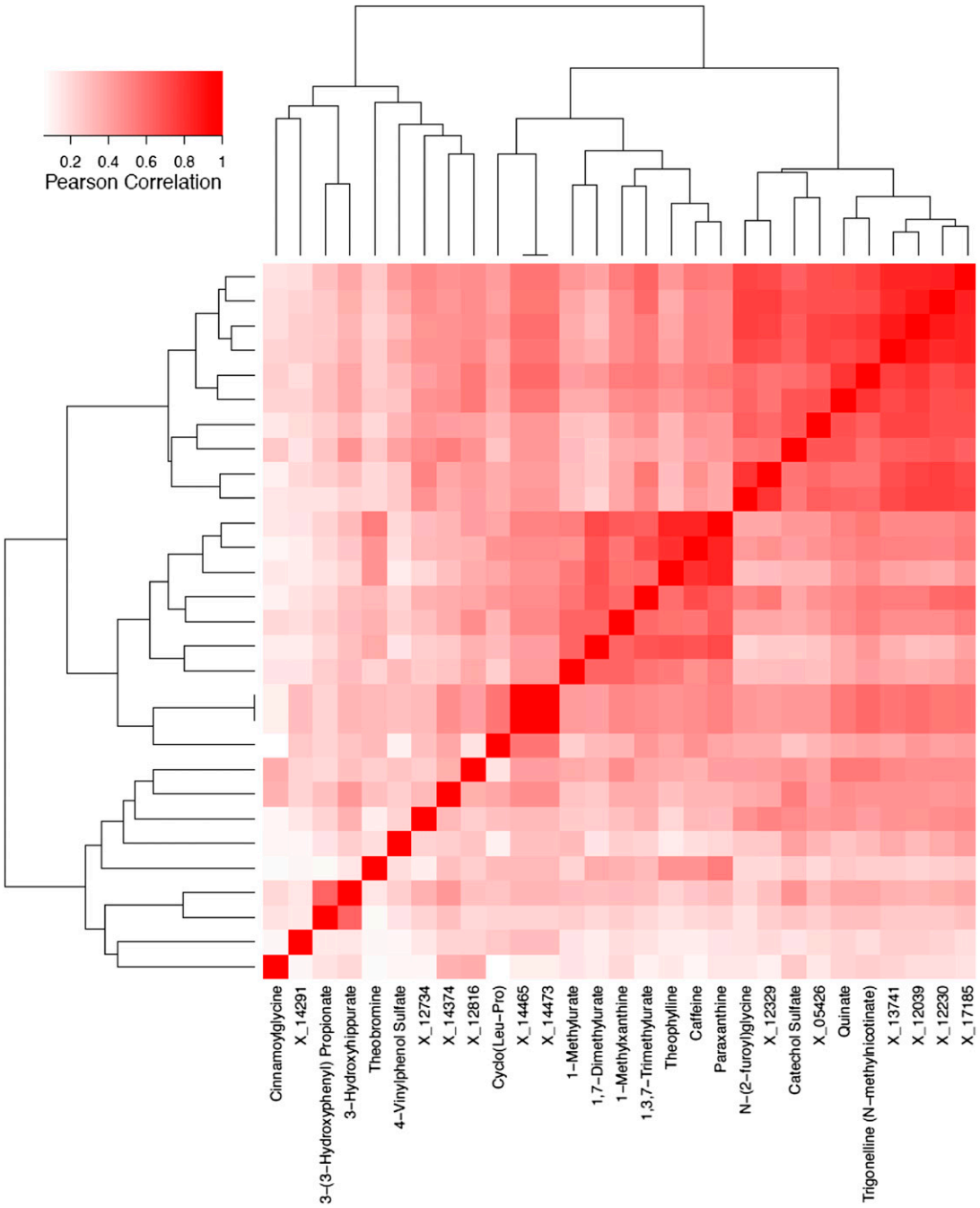

FIGURE 1 Metabolites with statistically significant coffee-metabolite associations ($P < 7.61 \times 10^{-5}$; Bonferroni-corrected threshold, $P = 0.05$ divided by the total number of metabolites detected) were highly intercorrelated.

previously reported in this cohort (39), the association between self-reported coffee intake and colorectal cancer was weakly inverse but NS. Although the self-reported recall of coffee consumption is one of the most accurately assessed aspects of diet by using FFQs (40), it can misclassify exposure because the actual metabolites to which an individual is exposed may vary considerably as a result of differences in coffee consumption (e.g., bean type, roast, and preparation method) (41–44) and individual metabolism (e.g., genotype and gut microbial composition) (21–23). Therefore, coffee biomarkers identified using untargeted metabolomic profiling may more accurately classify the exposure of participants to coffee compounds by serving as an integrated exposure measure (24).

Few previous studies have used untargeted metabolomic profiling to identify biomarkers of habitual coffee intake (45). The most comparable study, from Rothwell et al. (45), compared urine samples of high (n = 20) and low (n = 19) coffee consumers. Eleven of our coffee-associated serum metabolites matched (1-methylxanthine, 1,7-dimethyluate, 1-methylurate, trigonelline, paraxanthine, 3-hydroxyhippurate, and caffeine) or were closely related to [cyclo(ile-pro), dimethylxanthine glucuronide, trimethylurate, and 1,3- or 3,7-dimethylurate] the coffee-associated

TABLE 3
Serum metabolites associated with self-reported total coffee intake by type of coffee consumption compared with nonconsumers in a nested case-control study within the PLCO Cancer Screening Trial (n = 478)[1]

	Coffee-metabolite associations by type of coffee consumption					
	Caffeinated only compared with none (n = 257)		Decaffeinated only compared with none (n = 182)		Mixed compared with none (n = 183)	
Serum metabolite	Partial Pearson correlation	P	Partial Pearson correlation	P	Partial Pearson correlation	P
Trigonelline (N'-methylnicotinate)	0.593	5.54×10^{-46}*	0.426	3.76×10^{-22}*	0.523	1.81×10^{-34}*
Quinate	0.569	1.01×10^{-41}*	0.471	1.96×10^{-27}*	0.510	1.75×10^{-32}*
X_12039	0.493	3.16×10^{-30}*	0.357	1.23×10^{-15}*	0.415	4.78×10^{-21}*
X_13741	0.479	1.90×10^{-28}*	0.338	5.05×10^{-14}*	0.371	7.46×10^{-17}*
X_12816	0.456	1.56×10^{-25}*	0.289	1.70×10^{-10}*	0.422	8.28×10^{-22}*
X_14465	0.434	4.33×10^{-23}*	0.262	7.36×10^{-9}*	0.381	1.11×10^{-17}*
X_14473	0.434	4.33×10^{-23}*	0.262	7.36×10^{-9}*	0.381	1.11×10^{-17}*
X_17185	0.459	6.06×10^{-26}*	0.278	8.19×10^{-10}*	0.340	3.32×10^{-14}*
X_12230	0.448	1.35×10^{-24}*	0.298	4.26×10^{-11}*	0.363	3.74×10^{-16}*
Paraxanthine[2]	0.467	7.68×10^{-27}*	0.109	1.81×10^{-2}	0.354	2.59×10^{-15}*
N-(2-furoyl)glycine	0.338	4.95×10^{-14}*	0.236	2.26×10^{-7}*	0.271	2.35×10^{-9}*
Catechol sulfate	0.325	4.74×10^{-13}*	0.253	2.45×10^{-8}*	0.282	4.50×10^{-10}*
Caffeine[2]	0.401	1.38×10^{-19}*	0.057	2.16×10^{-1}	0.297	5.03×10^{-11}*
X_05426	0.343	2.01×10^{-14}*	0.213	3.08×10^{-6}*	0.272	2.10×10^{-9}*
1-Methylxanthine[2]	0.417	3.38×10^{-21}*	0.040	3.84×10^{-1}	0.308	7.71×10^{-12}*
Theophylline[2]	0.370	1.01×10^{-16}*	0.029	5.32×10^{-1}	0.303	1.81×10^{-11}*
X_12329	0.306	1.12×10^{-11}*	0.180	8.75×10^{-5}	0.235	2.37×10^{-7}*
1,3,7-Trimethylurate[2]	0.362	4.88×10^{-16}*	0.010	8.32×10^{-1}	0.225	8.23×10^{-7}*
3-Hydroxyhippurate	0.212	3.43×10^{-6}*	0.156	6.75×10^{-4}	0.198	1.52×10^{-5}*
1,7-Dimethylurate[2]	0.323	6.23×10^{-13}*	0.052	2.62×10^{-1}	0.196	1.80×10^{-5}*
1-Methylurate[2]	0.271	2.29×10^{-9}*	0.014	7.65×10^{-1}	0.181	7.80×10^{-5}
X_14291	0.183	6.53×10^{-5}*	0.122	8.15×10^{-3}	0.191	3.15×10^{-5}*
X_14374	0.211	3.70×10^{-6}*	0.132	4.04×10^{-3}	0.148	1.29×10^{-3}
Cyclo(leu-pro)	0.193	2.41×10^{-5}*	0.130	4.62×10^{-3}	0.139	2.47×10^{-3}
4-Vinylphenol sulfate	0.191	2.88×10^{-5}*	0.162	4.16×10^{-4}	0.139	2.54×10^{-3}
3-(3-Hydroxyphenyl) propionate	0.202	1.00×10^{-5}*	0.125	6.76×10^{-3}	0.150	1.13×10^{-3}
Theobromine[2]	0.179	9.13×10^{-5}*	0.059	2.01×10^{-1}	0.133	3.70×10^{-3}
X_12734	0.227	6.40×10^{-7}*	0.145	1.64×10^{-3}	0.175	1.36×10^{-4}
Cinnamoylglycine	0.184	6.00×10^{-5}*	0.095	3.98×10^{-2}	0.202	1.02×10^{-5}*

[1]Only significant metabolites from the main analysis are shown; the type of coffee intake was self-reported during follow-up on the Dietary Health Questionnaire; individuals with missing types of coffee intake (n = 20) were excluded. Each group of comparisons included n = 47 nondrinkers. Metabolites with the prefix X_ are of unknown identity. Partial correlations were determined between ln(metabolite) and relevant ln(coffee) (grams per day of caffeinated, decaffeinated, or mixed) adjusted for sex, tobacco smoking status (current, former, never), age (continuous), and BMI (continuous). P values are from the linear regression of relevant ln(coffee) (grams per of caffeinated, decaffeinated, or mixed) on ln(metabolite) adjusted for sex, tobacco smoking status (current, former, and never), age (continuous), and BMI (continuous). *Significant associations from the linear regression of ln(metabolite) on ln(coffee). Bonferroni-corrected level of significance, $P < 7.61 \times 10^{-5}$. PLCO, Prostate, Lung, Colorectal, and Ovarian.

[2]Caffeine and known caffeine metabolites.

urinary metabolites identified in this study. In addition, this case-control study nested within the PLCO Cancer Screening Trial was previously used to explore the cross-sectional associations of metabolites with usual dietary intake (25); Guertin et al. (25) showed that some of the strongest dietary-metabolite correlates were for coffee; these metabolites included trigonelline, quinate, 1-methylxanthine, paraxanthine, *N*-2-(furoyl)glycine, and catechol sulfate. In the current analysis, we focused on coffee-related metabolites with the a priori hypothesis that coffee-associated metabolites are inversely related to risk of colorectal cancer. Furthermore, we expanded on previous findings by investigating metabolites of both known and unknown identities.

Of 29 metabolites associated with self-reported coffee intake in our study, all 17 compounds of known identity could be plausibly linked to coffee (**Figure 2**). The human metabolism of caffeine, which is naturally present in green coffee beans and largely unaffected by roasting, has been well documented (21, 46, 47); caffeine and 7 caffeine metabolites (highlighted in blue in Figure 2) were significantly associated with coffee intake in our study. Three metabolites (catechol sulfate, 3-hydroxyhippurate, and 4-vinylphenol sulfate; highlighted in red in Figure 2) are related to benzoate metabolism. Benzoate is naturally present in coffee and also a downstream metabolite of chlorogenic acids, which are largely metabolized by microbial enzymes in the lower gastrointestinal tract (23, 48–50). Catechol, which is a degradation product of benzoate (46, 51), is formed from naturally occurring chlorogenic acids during coffee roasting and is conjugated to sulfate in plasma (52). The related metabolite 3-hydroxhippurate contains a benzoyl group and is a downstream metabolite of 5-*O*-caffeoylquinic acid (53), which is the most abundant chlorogenic

TABLE 4
ORs (95% CIs) for coffee-associated metabolites and diagnosis of colorectal cancer in a nested study within the PLCO Cancer Screening Trial (n = 251 incident cases and n = 247 controls) by using conditional logistic regression for metabolites with <20% of individuals below the lower limit of detection[1]

	OR (95% CI)			
	High metabolite levels compared with none/low		90th compared with 10th percentiles for each continuously measured metabolite	
Name	Undetectable/low	High		*P*-linear trend
Theophylline	Reference	0.75 (0.51, 1.09)	0.44 (0.25, 0.79)	0.006*
Cases/controls, *n*	147/132	104/115	—	
Caffeine	Reference	0.81 (0.55, 1.20)	0.56 (0.35, 0.89)	0.015*
Cases/controls, *n*	133/125	118/122	—	
Paraxanthine	Reference	0.93 (0.63, 1.36)	0.58 (0.36, 0.94)	0.027*
Cases/controls, *n*	133/130	118/117	—	
X_05426	Reference	0.83 (0.58, 1.20)	0.66 (0.41, 1.05)	0.080
Cases/controls, *n*	134/125	117/122	—	
X_14374	Reference	0.85 (0.59, 1.22)	0.82 (0.53, 1.26)	0.366
Cases/controls, *n*	133/123	118/124	—	
Theobromine	Reference	0.94 (0.65, 1.34)	0.83 (0.54, 1.28)	0.405
Cases/controls, *n*	133/126	118/121	—	
Catechol sulfate	Reference	0.98 (0.66, 1.43)	0.83 (0.49, 1.39)	0.473
Cases/controls, *n*	125/125	126/122	—	
Quinate	Reference	0.90 (0.61, 1.33)	0.83 (0.48, 1.44)	0.508
Cases/controls, *n*	150/146	101/101	—	
4-Vinylphenol sulfate	Reference	0.92 (0.63, 1.34)	1.09 (0.63, 1.87)	0.759
Cases/controls, *n*	134/132	117/115	—	

[1]Conditional logistic regression of case/control status on categorized relative level of metabolite; ln(metabolite) was rank ordered into low compared with high categories by the median (varied by metabolite). All models were adjusted for tobacco smoking status (current, former, and never), age, and BMI (continuous). Cases and controls were incidence-density matched to cases on age (5-y intervals), year of randomization, season of blood draw, sex, and race. Metabolites with the prefix X_ are of unknown identity. ORs compared subjects with levels below the median value (none/low) to those with levels above the median (high). To have an adequately sized reference group, in these models, individuals with metabolite levels below the lower limit of detection are included in the reference category of those below the median value; none refers to relative metabolite levels below the lower limit of detection. Metabolites with statistically significant coffee-metabolite associations are shown ($P < 7.61 \times 10^{-5}$; Bonferroni-corrected threshold, $P = 0.05$ divided by the total number of metabolites detected). Metabolites are ordered by *P*-linear trend. *P*-linear trend values are from a conditional logistic regression of case/control status on ln(metabolite). *Significant associations. PLCO, Prostate, Lung, Colorectal, and Ovarian.

acid in coffee (44). Finally, 4-vinylphenol sulfate is the sulfate-bound form of a vinyl derivative of the coffee constituent 4-hydroxycinnamic acid (i.e., *p*-coumarate) that is produced by some lactic acid bacteria (i.e., *Lactobacillus*) (54, 55). Whether 4-vinylphenol is formed endogenously from *p*-coumarate or in coffee beans and then ingested is unclear.

Metabolites highlighted in yellow and gray in Figure 2 are compounds or metabolites of compounds that occur naturally in green coffee beans or are produced by roasting. Quinate (highlighted in yellow in Figure 2) may be ingested or formed from 5-*O*-caffeoylquinic acid by gut bacteria (23, 48, 53). Similarly, 3-(3-hydroxyphenyl) propionate (highlighted in orange in Figure 2) is likely formed from 5-*O*-caffeoylquinic acid by gut bacteria (23, 48, 53). Cinnamoylglycine is the glycine conjugate of cinnamic acid; it is unknown whether cinnamoylglycine is formed endogenously from plant cinnamate or excreted in its ingested form (56). The furan-derived precursors of *N*-(2-furoyl)glycine and cyclo(leu-pro) are probable byproducts of coffee roasting (57, 58). Trigonelline is an alkaloid in green coffee beans that is degraded to some extent during roasting to produce niacin (i.e., vitamin B-3) (59), which has been shown to suppress colonic inflammation and carcinogenesis in mice (60).

Caffeine, which is metabolized in the liver by the cytochrome P450 oxidase enzyme system into theophylline and paraxanthine (21, 47), may directly reduce colorectal cancer risk. For example, studies showed that caffeinated coffee stimulates colonic motor activity to a greater extent than does decaffeinated coffee (61), caffeine inhibits colon cancer cell growth (62), and theophylline induces apoptosis in cancer cell lines (63). Alternatively, caffeine and theophylline may more closely approximate exposure to bioactive compounds in coffee that are involved in the cause of colorectal cancer. For example, of the 2 predominantly consumed coffee species *Coffea canephora* (i.e., robusta) and *Coffea arabica* (i.e., arabica), the former has a higher concentration of caffeine, caffeic acid, and soluble fiber than the latter (44). Caffeic acid, which has been shown to high antioxidant capacity in cells peaks in human plasma within 1 h of coffee consumption and is mainly current in glucuronate- and sulfate-bound forms (65). Consequently, future metabolomics studies of coffee intake and colorectal cancer should consider other biofluids including urine (66) and feces. A third possibility is that serum caffeine and theophylline levels are associated with an extraneous factor that is also associated with decreased risk of colorectal cancer. Finally, we could not discount chance as a possible explanation for our findings.

TABLE 5
ORs (95% CIs) for coffee-associated metabolites and diagnosis of colorectal cancer in a nested study within the PLCO Cancer Screening Trial (n = 251 incident cases and n = 247 controls) by using conditional logistic regression for metabolites with ≥20% of individuals below the lower limit of detection[1]

	OR (95% CI)				
	Low and high relative metabolite levels compared with none			90th compared with 10th percentiles for each continuously measured metabolite	
Name	Undetectable	Low	High		*P*-linear trend
1-Methylxanthine	Reference	0.81 (0.51, 1.28)	0.63 (0.39, 1.02)	0.65 (0.41, 1.05)	0.080
Cases/controls, *n*	105/89	76/76	70/82	—	
1,7-Dimethylurate	Reference	0.77 (0.49, 1.22)	0.76 (0.49, 1.19)	0.72 (0.46, 1.12)	0.140
Cases/controls, *n*	120/107	66/69	65/71	—	
Cyclo(leu-pro)	Reference	0.95 (0.60, 1.51)	0.70 (0.44, 1.09)	0.74 (0.50, 1.11)	0.146
Cases/controls, *n*	134/124	63/57	55/66	—	
N-(2-furoyl)glycine	Reference	0.68 (0.43, 1.08)	0.83 (0.52, 1.32)	0.77 (0.49, 1.24)	0.283
Cases/controls, *n*	110/99	66/78	75/70	—	
3-Hydroxyhippurate	Reference	0.87 (0.54, 1.40)	0.83 (0.50, 1.37)	0.79 (0.50, 1.24)	0.297
Cases/controls, *n*	149/143	51/52	51/52	—	
X_14291	Reference	1.04 (0.62, 1.73)	0.84 (0.50, 1.41)	0.78 (0.46, 1.33)	0.367
Cases/controls, *n*	51/48	102/97	98/102	—	
1,3,7-Trimethylurate	Reference	0.83 (0.51, 1.36)	0.83 (0.51, 1.35)	82 (0.52, 1.28)	0.380
Cases/controls, *n*	145/139	53/54	53/54	—	
X_13741	Reference	0.88 (0.58, 1.37)	0.90 (0.57, 1.44)	0.85 (0.53, 1.36)	0.489
Cases/controls, *n*	84/85	81/83	86/79	—	
X_12816	Reference	1.13 (0.72, 1.78)	0.89 (0.57, 1.41)	0.85 (0.54, 1.36)	0.499
Cases/controls, *n*	115/116	73/60	63/71	—	
Trigonelline (*N′*-methylnicotinate)	Reference	0.78 (0.48, 1.28)	0.77 (0.46, 1.23)	0.85 (0.50, 1.45)	0.558
Cases/controls, *n*	58/53	94/99	99/95	—	
Cinnamoylglycine	Reference	0.81 (0.48, 1.34)	0.85 (0.51, 1.42)	0.89 (0.60, 1.32)	0.561
Cases/controls, *n*	174/161	37/44	40/42	—	
X_12230	Reference	0.95 (0.61, 1.48)	0.92 (0.58, 1.45)	0.87 (0.54, 1.39)	0.561
Cases/controls, *n*	96/99	78/73	77/75	—	
X_14465	Reference	1.05 (0.63, 1.74)	0.92 (0.54, 1.56)	0.86 (0.50, 1.50)	0.601
Cases/controls, *n*	48/51	102/97	101/99	—	
X_14473	Reference	1.05 (0.63, 1.74)	0.92 (0.54, 1.56)	0.86 (0.50, 1.50)	0.601
Cases/controls, *n*	48/51	102/97	101/99	—	
3-(3-Hydroxyphenyl) propionate	Reference	1.45 (0.67, 3.17)	0.81 (0.41, 1.60)	0.90 (0.60, 1.35)	0.608
Cases/controls, *n*	214/216	19/11	18/20	—	
X_17185	Reference	1.18 (0.75, 1.86)	0.97 (0.61, 1.56)	0.91 (0.56, 1.49)	0.704
Cases/controls, *n*	95/104	79/70	77/73	—	
X_12329	Reference	0.98 (0.60, 1.61)	1.02 (0.62, 1.68)	0.93 (0.61, 1.44)	0.754
Cases/controls, *n*	159/162	45/43	47/42	—	
1-Methylurate	Reference	1.28 (0.80, 2.05)	0.84 (0.52, 1.35)	0.94 (0.59, 1.49)	0.791
Cases/controls, *n*	130/134	65/52	56/61	—	
X_12734	Reference	0.96 (0.61, 1.49)	0.96 (0.60, 1.53)	1.02 (0.67, 1.55)	0.927
Cases/controls, *n*	126/128	64/58	61/61	—	
X_12039	Reference	1.37 (0.87, 2.15)	0.86 (0.53, 1.39)	0.99 (0.61, 1.59)	0.963
Cases/controls, *n*	82/93	93/68	76/86	—	

[1]Conditional logistic regression of case/control status on categorized relative level of metabolite; ln(metabolite) was rank ordered into low compared with high categories by the median (varied by metabolite). ORs compared subjects with relative metabolite levels below the lower limit of detection (none) to those with detectable relative levels below (low) or above (high) the median relative level. All models were adjusted for tobacco smoking status (current, former, and never), age, and BMI (continuous). Cases and controls were incidence-density matched to cases on age (5-y intervals), year of randomization, season of blood draw, sex, and race. Metabolites with the prefix X_ are of unknown identity. None refers to relative metabolite levels below the lower limit of detection (undetectable). Metabolites with statistically significant coffee-metabolite associations are shown ($P < 7.61 \times 10^{-5}$; Bonferroni-corrected threshold, $P = 0.05$ divided by the total number of metabolites detected). Metabolites are ordered by *P*-linear trend. *P*-linear trend values are from a conditional logistic regression of case/control status on ln(metabolite). PLCO, Prostate, Lung, Colorectal, and Ovarian.

Our study had several limitations. The metabolomic platform we used was more comprehensive than those used by previous epidemiologic studies of coffee metabolites (35, 67), but many compounds remain unknown. Future work should include the identification of metabolites of unknown identity that were associated with coffee herein and in other studies (45). In addition, important chemopreventive compounds may be outside the scope of metabolomics (e.g., soluble fiber) or not easily measured in fasting serum because of their relatively short half-lives (e.g., caffeic acid). Because our primary hypothesis was that coffee metabolites are associated with colorectal cancer, we did not consider other potential sources of caffeine. In our analytic

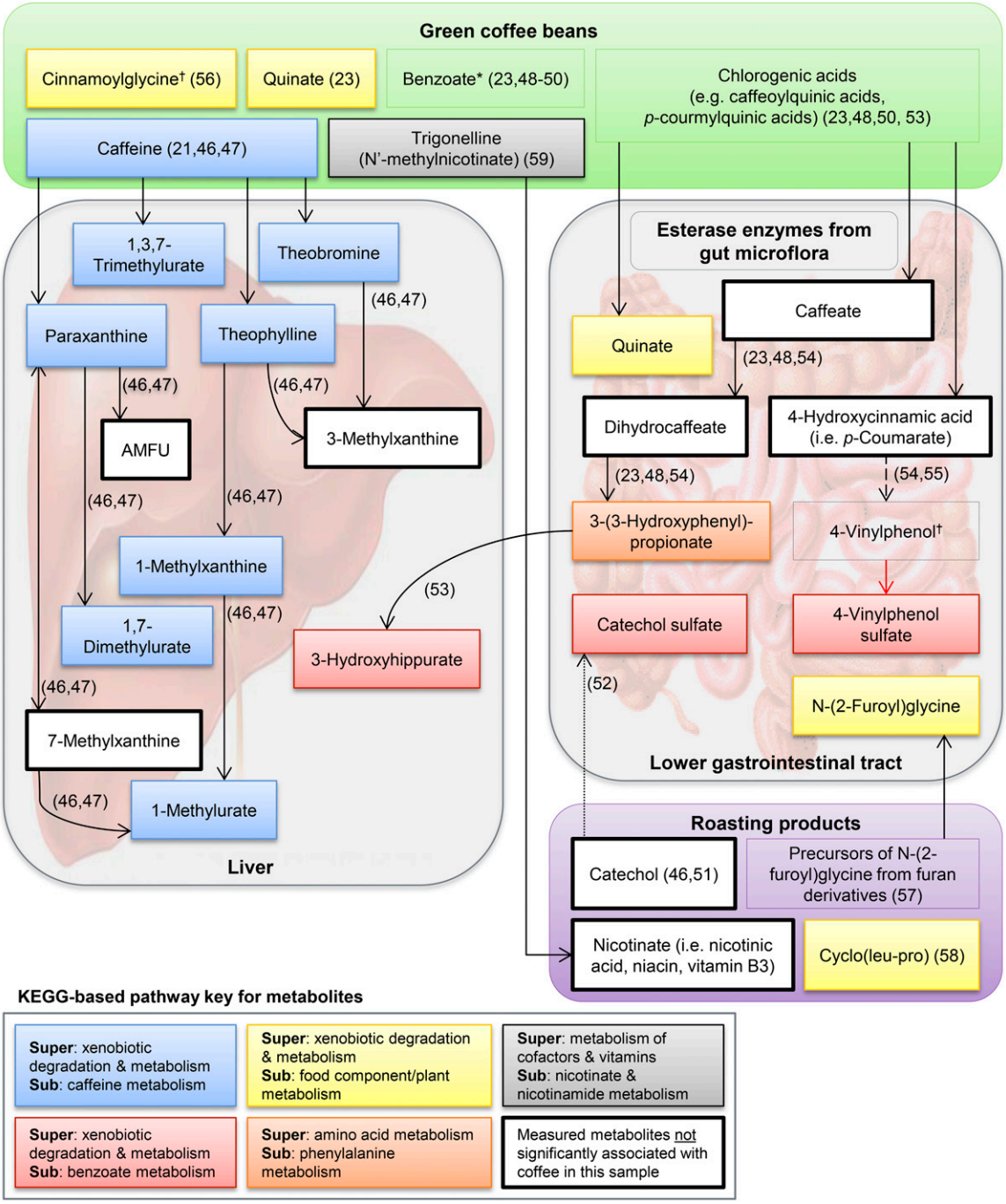

FIGURE 2 Proposed relations between the top named coffee-associated metabolites and green/roasted coffee on the basis of searches of the published literature, the Human Metabolome Database, and the Kyoto Encyclopedia of Genes and Genomes. The red arrow denotes that sulfonation catalyzed by a supergene family of enzymes called sulfotransferases and may occur in the cytosol of organs other than the intestine (64). The relation shown by the dashed arrow has not been observed in humans. *Multiple sources of benzoate related to coffee compounds and metabolites; †origin unclear. AMFU, 5-acetylamino-6-formylamino-3-methyluracil; KEGG, Kyoto Encyclopedia of Genes and Genomes.

sample, serum caffeine was highly correlated with coffee but not with tea, soda, or chocolate; partial Pearson correlations adjusted for age, sex, tobacco-smoking status, and BMI were 0.33 ($P = 1.02 \times 10^{-13}$), 0.04 ($P = 0.39$), −0.01 ($P = 0.74$), and −0.03 ($P = 0.49$), respectively. Although coffee was the primary source of caffeine in our sample, future studies may want to explore the association of total caffeine consumption with colorectal cancer. Our study was also limited by characteristics of the study sample, who were primarily (>90%) Caucasian. Consequently, the generalizability of our findings to other groups may be limited. Detailed information on self-reported coffee consumption, including the bean type, roast, and preparation method, was not captured by the available dietary questionnaires. Thus, we were unable to delineate differences in coffee-associated metabolites by differences in coffee beverages beyond caffeine status. Future studies should explore biomarkers of coffee intake

by using an untargeted metabolomic approach in conjunction with a more detailed coffee questionnaire. Reverse causality should be considered because individuals with no or low coffee consumption may have altered consumption as a result of early symptoms of disease (e.g., stomach pain); to mitigate the influence of reverse causality, we excluded cases who were diagnosed ≤6 mo of baseline. In addition, we examined the associations between metabolites and colorectal cancer by the median follow up time; results were similar, albeit generally nonsignificant because of smaller sample sizes, which suggested that reverse causality did not explain our findings. Finally, our study had limited power to detect modest associations of metabolites with colorectal cancer.

In conclusion, we show that an untargeted metabolomic approach can be used to identify serum metabolites that are associated with coffee drinking. Although the replication of these results is needed, we provide evidence that some metabolites associated with coffee intake, particularly caffeinated coffee intake, are also inversely associated with colorectal cancer. Future epidemiologic studies that use untargeted metabolomics to understand mechanisms underlying diet-disease associations are warranted.

The authors' responsibilities were as follows—SCM, JNS, W-YH, NDF, AJC, and RS: designed the research; KAG, EL, SMB, and RS: conducted the research; SCM, JNS, W-YH, NDF, and AJC: provided essential materials; KAG, EL, SMB, SCM, JNS, QX, and XX: analyzed data; and KAG, EL, and RS: wrote the manuscript and had primary responsibility for the final content of the manuscript. None of the authors reported a conflict of interest related to the study.

REFERENCES

1. Farah A, Monteiro M, Donangelo C, Lafay S. Chlorogenic acids from green coffee extract are highly bioavailable in humans. J Nutr 2008; 138:2309–15.
2. Nkondjock A. Coffee consumption and the risk of cancer: an overview. Cancer Lett 2009;277:121–5.
3. Arab L. Epidemiologic evidence on coffee and cancer. Nutr Cancer 2010;62:271–83.
4. Je Y, Giovannucci E. Coffee consumption and total mortality: a meta-analysis of twenty prospective cohort studies. Br J Nutr 2014;111: 1162–73.
5. Malerba S, Turati F, Galeone C, Pelucchi C, Verga F, La Vecchia C, Tavani A. A meta-analysis of prospective studies of coffee consumption and mortality for all causes, cancers and cardiovascular diseases. Eur J Epidemiol 2013;28:527–39.
6. Crippa A, Discacciati A, Larsson SC, Wolk A, Orsini N. Coffee consumption and mortality from all causes, cardiovascular disease, and cancer: a dose-response meta-analysis. Am J Epidemiol 2014;180: 763–75.
7. Freedman ND, Park Y, Abnet CC, Hollenbeck AR, Sinha R. Association of coffee drinking with total and cause-specific mortality. N Engl J Med 2012;366:1891–904.
8. Liu R, Guo X, Park Y, Huang X, Sinha R, Freedman ND, Hollenbeck AR, Blair A, Chen H. Caffeine intake, smoking, and risk of Parkinson disease in men and women. Am J Epidemiol 2012;175:1200–7.
9. Je Y, Giovannucci E. Coffee consumption and risk of endometrial cancer: findings from a large up-to-date meta-analysis. Int J Cancer 2012;131:1700–10.
10. Gunter MJ, Schaub JA, Xue X, Freedman ND, Gaudet MM, Rohan TE, Hollenbeck AR, Sinha R. A prospective investigation of coffee drinking and endometrial cancer incidence. Int J Cancer 2012;131: E530–6.
11. Lai GY, Weinstein SJ, Albanes D, Taylor PR, McGlynn KA, Virtamo J, Sinha R, Freedman ND. The association of coffee intake with liver cancer incidence and chronic liver disease mortality in male smokers. Br J Cancer 2013;109:1344–51.
12. Sang LX, Chang B, Li XH, Jiang M. Consumption of coffee associated with reduced risk of liver cancer: a meta-analysis. BMC Gastroenterol 2013;13:34.
13. Ferrucci LM, Cartmel B, Molinaro AM, Leffell DJ, Bale AE, Mayne ST. Tea, coffee, and caffeine and early-onset basal cell carcinoma in a case-control study. Eur J Cancer Prev 2014;23:296–302.
14. Song F, Qureshi AA, Han J. Increased caffeine intake is associated with reduced risk of basal cell carcinoma of the skin. Cancer Res 2012;72:3282–9.
15. Sinha R, Cross AJ, Daniel CR, Graubard BI, Wu JW, Hollenbeck AR, Gunter MJ, Park Y, Freedman ND. Caffeinated and decaffeinated coffee and tea intakes and risk of colorectal cancer in a large prospective study. Am J Clin Nutr 2012;96:374–81.
16. Howlader N, Noone AM, Krapcho M, Garshell J, Neyman N, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, et al., editors. SEER Cancer Statistics Review, 1975-2010. Version current November 2012 [cited 2014 Nov 4]. Available from: http://seer.cancer.gov/csr/1975_2010/.
17. Galeone C, Turati F, La Vecchia C, Tavani A. Coffee consumption and risk of colorectal cancer: a meta-analysis of case-control studies. Cancer Causes Control 2010;21:1949–59.
18. Je Y, Liu W, Giovannucci E. Coffee consumption and risk of colorectal cancer: a systematic review and meta-analysis of prospective cohort studies. Int J Cancer 2009;124:1662–8.
19. Yu X, Bao Z, Zou J, Dong J. Coffee consumption and risk of cancers: a meta-analysis of cohort studies. BMC Cancer 2011;11:96.
20. Zhang X, Albanes D, Beeson WL, van den Brandt PA, Buring JE, Flood A, Freudenheim JL, Giovannucci EL, Goldbohm RA, Jaceldo-Siegl K, et al. Risk of colon cancer and coffee, tea, and sugar-sweetened soft drink intake: pooled analysis of prospective cohort studies. J Natl Cancer Inst 2010;102:771–83.
21. Kot M, Daniel WA. The relative contribution of human cytochrome P450 isoforms to the four caffeine oxidation pathways: an in vitro comparative study with cDNA-expressed P450s including CYP2C isoforms. Biochem Pharmacol 2008;76:543–51.
22. Cornelis MC, Byrne EM, Esko T, Nalls MA, Ganna A, Paynter N, Monda KL, Amin N, Fischer K, Renstrom F, et al. Genome-wide meta-analysis identifies six novel loci associated with habitual coffee consumption. Mol Psychiatry 2014 Oct 7 (Epub ahead of print; DOI: 10.1038/mp.2014.107).
23. Gonthier MP, Remesy C, Scalbert A, Cheynier V, Souquet JM, Poutanen K, Aura AM. Microbial metabolism of caffeic acid and its esters chlorogenic and caftaric acids by human faecal microbiota in vitro. Biomed Pharmacother 2006;60:536–40.
24. Oresic M. Metabolomics, a novel tool for studies of nutrition, metabolism and lipid dysfunction. Nutr Metab Cardiovasc Dis 2009;19:816–24.
25. Guertin KA, Moore SC, Sampson JN, Huang WY, Xiao Q, Stolzenberg-Solomon RZ, Sinha R, Cross AJ. Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. Am J Clin Nutr 2014;100:208–17.
26. Cross AJ, Moore SC, Boca S, Huang WY, Xiong X, Stolzenberg-Solomon R, Sinha R, Sampson JN. A prospective study of serum metabolites and colorectal cancer risk. Cancer 2014;120:3049–57.
27. Gohagan JK, Prorok PC, Hayes RB, Kramer BS. The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial of the National Cancer Institute: history, organization, and status. Control Clin Trials 2000;21(6 Suppl):251S–72S.
28. Prorok PC, Andriole GL, Bresalier RS, Buys SS, Chia D, Crawford ED, Fogel R, Gelmann EP, Gilbert F, Hasson MA, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. Control Clin Trials 2000;21(6 Suppl):273S–309S.
29. World Health Organization. International Classification of Diseases for Oncology. 3rd ed (ICD-O-3). 2000 [cited 2014 Nov 4]. Available from: http://www.who.int/classifications/icd/adaptations/oncology/en/.
30. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH, Speizer FE. Reproducibility and validity of a semiquantitative food frequency questionnaire. Am J Epidemiol 1985;122: 51–65.
31. Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. A data-based approach to diet questionnaire design and testing. Am J Epidemiol 1986;124:453–69.
32. Tippett KS, Cypell YS. Design and operation: the continuing survey of food intakes by individuals and the diet and health knowledge survey, 1994-96. Nationwide Food Surveys Report No. 96-1. Springfield (VA): US Department of Agriculture, Agricultural Research Service; 1997. p. 1–139.

33. Subar AF, Thompson FE, Kipnis V, Midthune D, Hurwitz P, McNutt S, McIntosh A, Rosenfeld S. Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires: the Eating at America's Table Study. Am J Epidemiol 2001;154:1089–99.
34. Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. Anal Chem 2009;81:6656–67.
35. Suhre K, Shin SY, Petersen AK, Mohney RP, Meredith D, Wagele B, Altmaier E, CARDIoGRAM, Deloukas P, Erdmann J, et al. Human metabolic individuality in biomedical and pharmaceutical research. Nature 2011;477:54–60.
36. Sha W, da Costa KA, Fischer LM, Milburn MV, Lawton KA, Berger A, Jia W, Zeisel SH. Metabolomic profiling can predict which humans will develop liver dysfunction when deprived of dietary choline. FASEB J 2010;24:2962–75.
37. Hayes RB, Reding D, Kopp W, Subar AF, Bhat N, Rothman N, Caporaso N, Ziegler RG, Johnson CC, Weissfeld JL, et al. Etiologic and early marker studies in the prostate, lung, colorectal and ovarian (PLCO) cancer screening trial. Control Clin Trials 2000;21(6 Suppl): 349S–55S.
38. Sampson JN, Boca SM, Shu XO, Stolzenberg-Solomon RZ, Matthews CE, Hsing AW, Tan YT, Ji BT, Chow WH, Cai Q, et al. Metabolomics in epidemiology: sources of variability in metabolite measurements and implications. Cancer Epidemiol Biomarkers Prev 2013;22:631–40.
39. Dominianni C, Huang WY, Berndt S, Hayes RB, Ahn J. Prospective study of the relationship between coffee and tea with colorectal cancer risk: the PLCO Cancer Screening Trial. Br J Cancer 2013;109:1352–9.
40. Grobbee DE, Rimm EB, Giovannucci E, Colditz G, Stampfer M, Willett W. Coffee, caffeine, and cardiovascular disease in men. N Engl J Med 1990;323:1026–32.
41. Gross G, Jaccaud E, Huggett AC. Analysis of the content of the diterpenes cafestol and kahweol in coffee brews. Food Chem Toxicol 1997;35:547–54.
42. Urgert R, van der Weg G, Kosmeijer-Schuil TG, van de Bovenkamp P, Hovenier R, Katan MB. Levels of the cholesterol-elevating diterpenes cafestol and kahweol in various coffee brews. J Agric Food Chem 1995;43:2167–72.
43. Alves RC, Almeida IM, Casal S, Oliveira MB. Isoflavones in coffee: influence of species, roast degree, and brewing method. J Agric Food Chem 2010;58:3002–7.
44. Farah A. Coffee constituents. In: Chu Y-F, editor. Coffee: emerging health effects and disease prevention. Oxford (United Kingdom): Wiley-Blackwell; 2012. p. 21–58.
45. Rothwell JA, Fillatre Y, Martin JF, 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.
46. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27–30.
47. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res 2014;42:D199–205.
48. Ludwig IA, Paz de Pena M, Concepcion C, Alan C. Catabolism of coffee chlorogenic acids by human colonic microbiota. Biofactors 2013;39:623–32.
49. Gonthier MP, Verny MA, Besson C, Remesy C, Scalbert A. Chlorogenic acid bioavailability largely depends on its metabolism by the gut microflora in rats. J Nutr 2003;133:1853–9.
50. Adamson RH, Bridges JW, Evans ME, Williams RT. Species differences in the aromatization of quinic acid in vivo and the role of gut bacteria. Biochem J 1970;116:437–43.
51. Reiner AM. Metabolism of aromatic compounds in bacteria. Purification and properties of the catechol-forming enzyme, 3,5-cyclohexadiene-1,2-diol-1-carboxylic acid (NAD +) oxidoreductase (decarboxylating). J Biol Chem 1972;247:4960–5.
52. Lang R, Dieminger N, Beusch A, Lee YM, Dunkel A, Suess B, Skurk T, Wahl A, Hauner H, Hofmann T. Bioappearance and pharmacokinetics of bioactives upon coffee consumption. Anal Bioanal Chem 2013;405:8487–503.
53. Rechner AR, Smith MA, Kuhnle G, Gibson GR, Debnam ES, Srai SK, Moore KP, Rice-Evans CA. Colonic metabolism of dietary polyphenols: influence of structure on microbial fermentation products. Free Radic Biol Med 2004;36:212–25.
54. van Beek S, Priest FG. Decarboxylation of substituted cinnamic acids by lactic acid bacteria isolated during malt whisky fermentation. Appl Environ Microbiol 2000;66:5322–8.
55. Rodriguez H, Landete JM, de las Rivas B, Munoz R. Metabolism of food phenolic acids by Lactobacillus plantarum CECT 748(T). Food Chem 2008;107:1393–8.
56. Hoskins JA, Holliday SB, Greenway AM. The metabolism of cinnamic acid by healthy and phenylketonuric adults: a kinetic study. Biomed Mass Spectrom 1984;11:296–300.
57. Pettersen JE, Jellum E. The identification and metabolic origin of 2-furoylglycine and 2,5-furandicarboxylic acid in human urine. Clin Chim Acta 1972;41:199–207.
58. Ginz M, Engelhardt UH. Identification of proline-based diketopiperazines in roasted coffee. J Agric Food Chem 2000;48:3528–32.
59. Allred KF, Yackley KM, Vanamala J, Allred CD. Trigonelline is a novel phytoestrogen in coffee beans. J Nutr 2009;139:1833–8.
60. Singh N, Gurav A, Sivaprakasam S, Brady E, Padia R, Shi H, Thangaraju M, Prasad PD, Manicassamy S, Munn DH, et al. Activation of Gpr109a, receptor for niacin and the commensal metabolite butyrate, suppresses colonic inflammation and carcinogenesis. Immunity 2014;40:128–39.
61. Rao SS, Welcher K, Zimmerman B, Stumbo P. Is coffee a colonic stimulant? Eur J Gastroenterol Hepatol 1998;10:113–8.
62. Merighi S, Benini A, Mirandola P, Gessi S, Varani K, Simioni C, Leung E, Maclennan S, Baraldi PG, Borea PA. Caffeine inhibits adenosine-induced accumulation of hypoxia-inducible factor-1alpha, vascular endothelial growth factor, and interleukin-8 expression in hypoxic human colon cancer cells. Mol Pharmacol 2007;72:395–406.
63. Wiernik PH, Paietta E, Goloubeva O, Lee SJ, Makower D, Bennett JM, Wade JL, Ghosh C, Kaminer LS, Pizzolo J, et al. Phase II study of theophylline in chronic lymphocytic leukemia: a study of the Eastern Cooperative Oncology Group (E4998). Leukemia 2004;18:1605–10.
64. Gamage N, Barnett A, Hempel N, Duggleby RG, Windmill KF, Martin JL, McManus ME. Human sulfotransferases and their role in chemical metabolism. Toxicol Sci 2006;90:5–22.
65. Nardini M, Cirillo E, Natella F, Scaccini C. Absorption of phenolic acids in humans after coffee consumption. J Agric Food Chem 2002; 50:5735–41.
66. Mennen LI, Sapinho D, Ito H, Bertrais S, 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.
67. Altmaier E, Kastenmüller G, Römisch-Margl W, Thorand B, Weinberger KM, Adamski J, Illig T, Döring A, Suhre K. Variation in the human lipidome associated with coffee consumption as revealed by quantitative targeted metabolomics. Mol Nutr Food Res 2009;53:1357–65.