



Metabolomic Signatures of Long-term Coffee Consumption and Risk of Type 2 Diabetes in Women

Diabetes Care 2020;43:2588–2596 | <https://doi.org/10.2337/dc20-0800>

Dong Hang,^{1,2} Oana A. Zeleznik,³ Xiaosheng He,^{4,5} Marta Guasch-Ferre,^{2,3} Xia Jiang,⁶ Jun Li,² Liming Liang,^{7,8} A. Heather Eliassen,^{3,7} Clary B. Clish,⁹ Andrew T. Chan,^{3,4,9} Zhibin Hu,¹ Hongbing Shen,¹ Kathryn M. Wilson,^{3,7} Lorelei A. Mucci,⁷ Qi Sun,^{2,3} Frank B. Hu,^{2,3,7} Walter C. Willett,^{2,3,7} Edward L. Giovannucci,^{2,3,7} and Mingyang Song^{2,4,7}

OBJECTIVE

Coffee may protect against multiple chronic diseases, particularly type 2 diabetes, but the mechanisms remain unclear.

RESEARCH DESIGN AND METHODS

Leveraging dietary and metabolomic data in two large cohorts of women (the Nurses' Health Study [NHS] and NHSII), we identified and validated plasma metabolites associated with coffee intake in 1,595 women. We then evaluated the prospective association of coffee-related metabolites with diabetes risk and the added predictivity of these metabolites for diabetes in two nested case-control studies ($n = 457$ case and 1,371 control subjects).

RESULTS

Of 461 metabolites, 34 were identified and validated to be associated with total coffee intake, including 13 positive associations (primarily trigonelline, polyphenol metabolites, and caffeine metabolites) and 21 inverse associations (primarily triacylglycerols [TAGs] and diacylglycerols [DAGs]). These associations were generally consistent for caffeinated and decaffeinated coffee, except for caffeine and its metabolites that were only associated with caffeinated coffee intake. The three cholesteryl esters positively associated with coffee intake showed inverse associations with diabetes risk, whereas the 12 metabolites negatively associated with coffee (5 DAGs and 7 TAGs) showed positive associations with diabetes. Adding the 15 diabetes-associated metabolites to a classical risk factor–based prediction model increased the C-statistic from 0.79 (95% CI 0.76, 0.83) to 0.83 (95% CI 0.80, 0.86) ($P < 0.001$). Similar improvement was observed in the validation set.

CONCLUSIONS

Coffee consumption is associated with widespread metabolic changes, among which lipid metabolites may be critical for the antidiabetes benefit of coffee. Coffee-related metabolites might help improve prediction of diabetes, but further validation studies are needed.

Coffee is one of the most popular beverages worldwide. Accumulating evidence indicates that long-term coffee intake is associated with lower risks of various chronic diseases, including type 2 diabetes, cardiovascular disease, and some types of cancer (1). Therefore, the 2015–2020 U.S. Dietary Guidelines recommend moderate coffee consumption as part of a healthy dietary pattern (2).

¹Department of Epidemiology and Biostatistics, Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Personalized Medicine, School of Public Health, Nanjing Medical University, Nanjing, China

²Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA

³Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

⁴Clinical and Translational Epidemiology Unit and Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA

⁵Department of Colorectal Surgery, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

⁶Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA

⁷Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA

⁸Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

⁹Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA

Corresponding author: Mingyang Song, mingyangsong@mail.harvard.edu

Received 10 April 2020 and accepted 12 July 2020

This article contains supplementary material online at <https://doi.org/10.2337/figshare.12675611>.

© 2020 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. More information is available at <https://www.diabetesjournals.org/content/license>.

Despite these data, the biological mechanisms underlying the benefit of coffee remain unclear. Coffee contains >1,000 bioactive compounds. We and others have shown that long-term coffee intake may reduce insulin resistance and inflammation and modulate hormonal levels that are key to cardiometabolic diseases and cancer (3,4). Recently, high-throughput metabolomic profiling has shown great promise in improving our understanding about the biological effects of nutritional factors and may help identify novel biomarkers for risk prediction and targets for intervention (5).

Several observational studies have examined the metabolomic profiles associated with coffee consumption (6–9). Although some metabolites have been commonly identified, such as trigonelline, caffeine metabolites, and quinate, findings for other metabolites have been inconsistent. This inconsistency may be partly due to limited coverage of the metabolomic platforms, small sample size, and insufficient control for confounding by lifestyle and other dietary factors. Moreover, few studies have investigated whether the identified metabolites influence chronic disease risk.

Therefore, we performed a systematic analysis of plasma metabolomics to 1) identify and validate metabolites associated with intake of total, caffeinated, and decaffeinated coffee and 2) prospectively evaluate the association of coffee-related metabolites with diabetes risk and the added predictivity of these metabolites for diabetes. We drew data from two large U.S. cohorts, the Nurses' Health Study (NHS) and NHSII.

RESEARCH DESIGN AND METHODS

Study Design and Population

The NHS enrolled 121,700 female nurses aged 30–55 years in 1976, and the NHSII enrolled 116,429 female nurses aged 25–42 years in 1989. Details about the two cohorts have previously been published (10). In brief, mailed questionnaires were administered biennially to assess lifestyle and medical history, with the follow-up rates exceeding 90% for each cycle in both cohorts. Blood samples were collected from 32,826 women in NHS in 1989–1990 and 29,611 women in NHSII in 1996–1999. Women who provided blood samples had dietary and lifestyle profiles similar to those of women who did not (11). Samples were returned by

overnight mail with an ice pack, and >95% of them arrived within 24 h of blood draw. Upon arrival, samples were immediately centrifuged and aliquoted into cryotubes as plasma, buffy coat, and red blood cells and stored in liquid nitrogen freezers.

The current study was designed in two phases. First, we performed a cross-sectional analysis to identify and validate coffee-related metabolites. We drew participants with available plasma metabolomic data from previous substudies in the cohorts. After exclusion of participants who had a history of diabetes, CVD, or cancer prior to blood draw or had missing data on coffee consumption or metabolites, a total of 949 women with 427 named metabolites and 646 women with 413 metabolites were included in the discovery and validation sets, respectively. Details about participant selection are provided in the Supplementary Materials and Supplementary Fig. 1.

Then we conducted a nested, 1:3 matched case-control study (207 case and 621 control subjects) to prospectively evaluate the metabolites associated with coffee intake in relation to risk of type 2 diabetes. We developed a diabetes prediction model using the identified metabolites and validated its performance in another case-control study (250 case and 750 control subjects). Details about the case and control subject selection are provided in the Supplementary Materials. Briefly, among participants with available metabolomic data, we identified incident cases of diabetes that occurred after blood draw until the end of follow-up (NHS, 30 June 2012, and NHSII, 30 June 2013). Using risk set sampling, we randomly selected three control subjects for each diabetes case subject among participants who were alive and free of diabetes at the time of diagnosis of the case subjects, matched on age at blood draw ± 1 year, study cohort (NHS, NHSII), and fasting status (Supplementary Fig. 1).

The study protocol was approved by the Institutional Review Board of the Brigham and Women's Hospital and the Human Subjects Committee Review Board of the Harvard T.H. Chan School of Public Health.

Dietary Assessment

A validated 131-item food-frequency questionnaire (FFQ) was administered every 4 years for collection of updated dietary

data. Participants were asked how often (ranging from “never or less than once per month” to “six or more times per day”), on average, they consumed a standard portion size of each food item during the previous year. For coffee, the questionnaire inquired about the frequency of consumption for caffeinated and decaffeinated coffee, separately, in an 8-oz (237-mL) cup. We calculated total coffee consumption as the sum of these two types of coffee. The validity and reproducibility of FFQs have previously been reported (12), with a high correlation between coffee consumption assessed by FFQs and four 1-week diet records ($r = 0.78$). In the current study, to capture regular dietary intake that is most relevant to metabolite levels and reduce within-person variability, we calculated the average of intake from the two FFQs administered most proximately to the time of blood draw (i.e., 1986 and 1990 for the NHS and 1995 and 1999 for NHSII). The distribution of coffee intake in each FFQ cycle has been shown to be highly stable in our cohorts (13).

Metabolomics Profiling

All the metabolomics data used in the study were generated at the Broad Institute using liquid chromatography–tandem mass spectrometry (LC-MS) as previously described (14,15). In brief, high-resolution, accurate mass profiling data were acquired using LC-MS systems comprised of Nexera X2 UHPLC systems (Shimadzu Corp., Marlborough, MA) coupled to a Q Exactive or Exactive Plus Orbitrap Mass Spectrometer (Thermo Fisher Scientific, Waltham, MA). Hydrophilic interaction liquid chromatography with positive-ion mode mass spectrometry detection was used to separate polar metabolites, while C18 chromatography with negative-ion mode detection and C8 chromatography with positive-ion mode detection were used to profile metabolites of intermediate polarity and lipids, respectively (see more details in Supplementary Materials). Raw data were processed using TraceFinder 3.3 software (Thermo Fisher Scientific) and Progenesis Q1 (Nonlinear Dynamics, Newcastle upon Tyne, U.K.). For each of the methods, metabolite identities were confirmed using authentic reference standards or reference samples. For assessment of temporal drift of the metabolomics profiling, pooled reference plasma samples

were interspersed and analyzed at intervals of ~20 participant samples. In each substudy, we removed unnamed metabolites, metabolites with no between-person variations, and metabolites with a mean coefficient of variation of >25% or an intraclass correlation coefficient of <0.40 among drift reference samples. Metabolites with <10% missingness were included, and missing data for each metabolite were imputed using half of the minimum measured value. Finally, a total of 427 metabolites were included in the analysis (Supplementary Table 1).

Statistical Analyses

For accounting for batch effect and improving normality, metabolite measurements were natural log transformed and standardized using z scores (SDs from the mean) within each substudy. In the phase I analysis, we used linear regression to examine the association of coffee intake with each of the metabolites after adjusting for potential confounders. (See details in Table 2 footnote. Details of covariate assessment are described in Supplementary Materials.) The results were presented as percentage difference in metabolite levels per 1-cup increment in coffee intake, using the following exponential function: $[\exp(\beta\text{-coefficient}) - 1] \times 100\%$ (3). We considered total coffee intake as our primary exposure and ran the analysis in the discovery and validation sets separately. In the secondary analysis, we examined caffeinated and decaffeinated coffee separately in the pooled samples. Sensitivity analyses were performed among never smokers to minimize any confounding by smoking and among participants who were selected as control subjects in the source case-control studies.

In the phase II analysis, we calculated odds ratios (ORs) and 95% CIs for the association between coffee-related metabolites (per 1-SD increase) and diabetes risk using conditional logistic regression. We adjusted for family history of diabetes in addition to other covariates included in the phase I analysis. We also performed principal component analysis (PCA) of coffee-related metabolites and examined the top two component scores in relation to diabetes risk.

To assess the predictivity of coffee-related metabolites for diabetes, we calculated the C-statistics using receiver operating characteristic analysis. We considered three models: model 1, based on

established risk factors, including age, BMI, physical activity, smoking, and family history of diabetes; model 2, based on the coffee-related metabolites that were associated with diabetes risk in the current study; and model 3, based on a combination of established factors and metabolites. A nonparametric method was used to compare the C-statistics between these models (16). Furthermore, we calculated the integrated discrimination improvement and the net reclassification improvement to evaluate the added predictive ability of metabolites (17).

All analyses were performed using R 3.2.5 (R Foundation for Statistical Computing, Vienna, Austria) and SAS, version 9.4 (SAS Institute, Cary, NC). All statistical tests were two sided. We calculated the false discovery rate (FDR) to correct for multiple testing at $\alpha = 0.05$ significance level (18).

RESULTS

Identification of Coffee-Related Metabolites

Table 1 shows basic characteristics of study participants according to frequency of total coffee consumption. Participants in the discovery ($n = 949$) and validation ($n = 646$) subsets had similar characteristics. Those who drank more coffee were more likely to be current smokers.

In the discovery phase, 73 of the 427 metabolites were associated with total coffee intake after multiple testing correction (FDR P value <0.05). Of the 73 metabolites, 10 were unavailable in the validation stage. Among the remaining 63 metabolites, 34 were validated to be associated with coffee intake (Table 2), of which 13 were positively associated with total coffee intake, including trigonelline, caffeine and its metabolites (5-acetylamino-6-amino-3-methyluracil [AAMU], 1,7-dimethyluric acid, and 7-methylxanthine), polyphenol metabolites (cinnamoylglycine and 4-hydroxyhippuric acid), phenyllactic acid, cytosine, L-carnitine, and three cholesteryl esters (CEs). The other 21 metabolites were inversely associated with total coffee intake, mostly triacylglycerols (TAGs) and diacylglycerols (DAGs). Correlation analysis of the 34 metabolites showed that most lipid species were positively correlated and metabolites of the same class or sharing fatty acid chains tended to cluster together (Supplementary Fig. 2).

Most of the 34 identified metabolites demonstrated consistent associations with

caffeinated and decaffeinated coffee, except for caffeine and its metabolites, which were associated with caffeinated coffee only (Table 2).

The coffee-metabolite associations remained essentially unchanged when we restricted the analysis to never smokers (Supplementary Table 2) or control participants of the original case-control studies (Supplementary Table 3).

Association of Coffee-Related Metabolites With Diabetes Risk

Basic characteristics of 457 diabetes case subjects and 1,371 matched control subjects (in the training and validation sets, respectively) at blood draw are shown in Supplementary Table 4. The multivariate-adjusted OR of diabetes per 1 cup/day increase in total coffee consumption was 0.95 (95% CI 0.82, 1.09) in the training set, 0.93 (95% CI 0.83, 1.03) in the validation set, and 0.95 (95% CI 0.89, 1.02) in the combined analysis.

Among the 34 coffee-related metabolites, 30 were available for the diabetes analysis. Cinnamoylglycine, phenyllactic acid, docosahexaenoic acid, and C52:5 TAG were unavailable due to high missingness (>50%). As shown in Table 3, after adjustment for potential confounders and correction for multiple testing, three metabolites positively associated with coffee intake showed an inverse association with diabetes risk (C18:1 CE, C20:4 CE, and C18:2 CE), with the adjusted ORs ranging from 0.59 to 0.61 per 1-SD increment. In contrast, 12 metabolites negatively associated with coffee showed a positive association with diabetes, including 5 DAGs and 7 TAGs, with the adjusted ORs ranging from 1.35 to 2.06 per 1-SD increment. Of the 15 metabolites associated with diabetes, 14 were available in the validation case-control set and showed consistent associations with diabetes (Supplementary Table 5).

The PCA analysis of the 30 coffee-related metabolites showed that the first three components explained 44.9%, 15.1%, and 10.3% of the total variation, respectively (Supplementary Figs. 3–5). The first principal component represented a mix of DAGs and TAGs and was positively associated with diabetes risk (OR 1.15 per 1-unit increase in the component score [95% CI 1.08, 1.22]), the second principal component mainly represented CEs and was inversely associated with diabetes

Table 1—Baseline characteristics of study participants in the NHS and NHSII according to total coffee consumption*

	Screening stage (n = 949)				Validation stage (n = 646)			
	Nondrinker	≤1 cup/ day	2–3 cups/ day	≥4 cups/ day	Nondrinker	≤1 cup/ day	2–3 cups/ day	≥4 cups/ day
Participants, n (%)	163 (17)	178 (19)	420 (44)	188 (20)	83 (13)	113 (17)	282 (44)	168 (26)
White (%)	98	97	98	100	99	97	98	97
Fasting blood (%)	67	75	67	55	70	74	69	70
Age at blood draw (years)	46.7 (7.5)	49.1 (8.1)	51.0 (8.0)	52.1 (7.2)	46.8 (7.4)	50.6 (7.4)	51.9 (7.7)	52.7 (7.8)
Total coffee (cups/day)†	0.0 (0.0)	0.5 (0.2)	2.1 (0.5)	4.4 (0.8)	0.0 (0.0)	0.6 (0.2)	2.2 (0.4)	4.3 (0.7)
Caffeinated coffee (cups/day)†	0.0 (0.0)	0.3 (0.3)	1.5 (0.8)	3.0 (1.1)	0.0 (0.0)	0.3 (0.2)	1.6 (0.8)	3.1 (1.1)
Decaffeinated coffee (cups/day)†	0.0 (0.0)	0.2 (0.2)	0.6 (0.7)	1.4 (1.1)	0.0 (0.0)	0.3 (0.2)	0.6 (0.6)	1.2 (1.0)
BMI (kg/m ²)†	25.5 (3.8)	25.6 (3.8)	25.0 (3.8)	25.4 (3.4)	26.5 (2.8)	25.4 (2.8)	25.1 (3.8)	24.8 (2.6)
Physical activity (MET h/week)†	23.2 (14.5)	19.0 (13.1)	19.7 (15.9)	18.7 (11.3)	16.7 (9.7)	19.2 (13.4)	18.2 (17.0)	16.1 (9.4)
Smoking status (%)								
Never	83	71	54	39	75	62	44	39
Past	15	25	38	48	19	34	46	42
Current	2	5	9	13	6	4	10	19
Alcohol consumption (g/day)†	2.5 (4.3)	3.4 (3.8)	6.7 (7.9)	5.4 (5.1)	3.2 (3.3)	2.8 (3.1)	6.0 (6.8)	6.9 (6.8)
Total energy intake (kcal/day)†	1,786 (293)	1,862 (301)	1,816 (410)	1,847 (340)	1,854 (227)	1,833 (244)	1,785 (392)	1,818 (290)
Hypertension (%)	17	15	18	15	13	15	12	15
Hypercholesterolemia (%)	29	39	30	33	26	43	32	34
Postmenopausal (%)	25	32	41	40	30	52	54	55
Current menopausal hormone use (%)‡	41	54	51	57	93	55	61	56
Regular multivitamin use (%)	53	49	50	46	49	41	48	41
Regular aspirin/NSAID use (%)§	46	46	52	58	54	50	59	73

Mean (SD) is presented for continuous variables and percentage for categorical variables. *All variables are standardized by age at blood draw except age. †Cumulative average assessments most proximate to blood draw were used to derive the values. ‡Defined in menopausal women only. §Regular users are defined as those taking ≥2 tablets of aspirin (325 mg/tablet) or nonsteroidal anti-inflammatory drugs (NSAIDs) per week.

risk (OR 0.80 [95% CI 0.71, 0.89]), and the third principal component represented caffeine and its metabolites and was not associated with diabetes (OR 0.96 [95% CI 0.86, 1.08]) (Table 3).

For the diabetes prediction analysis (Fig. 1), we obtained similar C-statistics for the model using the 15 diabetes-associated metabolites and that using the classical risk factors. Adding metabolites to the classical risk factor model increased the C-statistic from 0.79 (95% CI 0.76, 0.83) to 0.83 (95% CI 0.80, 0.86) ($P < 0.001$) in the training set and from 0.75 (95% CI 0.72, 0.78) to 0.78 (95% CI 0.75, 0.82) ($P < 0.001$) in the validation set. The C-statistics were essentially unchanged after addition of coffee consumption to the models (data not shown). Among individual metabolites, C50:2 TAG and C34:1 DAG showed the highest C-statistics in the training set (both 0.74) and validation set (0.68 and 0.69, respectively) (Supplementary Table 6).

Compared with the models with only classical risk factors, statistically significant ($P < 0.0001$) higher net reclassification improvement was observed for

the model combining classical factors and metabolites (training set, 62.2% [95% CI 47.2, 77.1], and validation set, 47.7% [95% CI 33.8, 61.7]) (Supplementary Table 7). Similarly, we found statistically significant improvement for the integrated discrimination improvement ($P < 0.0001$) for the model combining classical factors and metabolites.

CONCLUSIONS

Leveraging the integrated dietary and metabolomic data in two large cohorts of women, we identified and validated 34 plasma metabolites associated with coffee intake. These metabolites can be divided into two groups: internal exposure markers of coffee intake (trigonelline, polyphenol metabolites, and caffeine and its metabolites) and metabolomic response to long-term coffee consumption such as lipid metabolites. Linking these metabolites to diabetes risk, we found that three CEs positively associated with coffee intake were associated with lower risk of diabetes, whereas 12 lipid metabolites (DAGs and TAGs) negatively associated with coffee were associated

with higher risk of diabetes. Moreover, the receiver operating characteristic analysis indicated that coffee-related metabolites might be useful to improve the prediction of diabetes beyond established risk factors. These findings provide new insights into the health benefit of coffee and suggest the potential of coffee-related metabolites for improved diabetes prediction.

A meta-analysis of 18 prospective cohort studies showed 7% lower risk of diabetes per 1 cup of coffee consumed per day (19). In addition, our previous findings in the NHS and NHSII support a beneficial role of both caffeinated and decaffeinated coffee in diabetes (20,21). In the current study, we found a similar inverse association between coffee intake and diabetes risk, although the 95% CIs contained one, likely due to the modest sample size. Also, we found that higher concentrations of trigonelline, an established marker of coffee exposure, were associated with lower risk of diabetes (OR per 1 SD 0.88 [95% CI 0.77, 0.99]) in the combined training and validation sets.

Table 3—Association between coffee-related metabolites and risk of type 2 diabetes in the NHS and NHSII*

HMDB ID	Metabolite	Super class	207 case vs. 621 control subjects		
			OR per 1 SD (95% CI)	P	FDR
Positively associated with coffee					
HMDB00610	C18:2 CE	Lipids and lipid-like molecules	0.59 (0.48, 0.72)	<0.0001	<0.001
HMDB00918	C18:1 CE	Lipids and lipid-like molecules	0.60 (0.50, 0.73)	<0.0001	<0.001
HMDB06726	C20:4 CE	Lipids and lipid-like molecules	0.61 (0.50, 0.75)	<0.0001	<0.001
HMDB00875	Trigonelline	Alkaloids and derivatives	0.87 (0.70, 1.08)	0.22	0.28
HMDB00630	Cytosine	Organoheterocyclic compounds	0.93 (0.77, 1.13)	0.46	0.49
HMDB04400	AAMU	Organic nitrogen compounds	1.04 (0.82, 1.33)	0.74	0.76
HMDB01991	7-methylxanthine	Organoheterocyclic compounds	1.09 (0.87, 1.35)	0.46	0.49
HMDB13678	4-hydroxyhippuric acid	Benzenoids	1.12 (0.92, 1.37)	0.26	0.33
HMDB01847	Caffeine	Organoheterocyclic compounds	1.18 (0.93, 1.49)	0.17	0.27
HMDB11103	1,7-dimethyluric acid	Organoheterocyclic compounds	1.22 (0.96, 1.56)	0.11	0.19
HMDB00062	L-carnitine	Organic nitrogen compounds	1.25 (0.99, 1.57)	0.06	0.12
Negatively associated with coffee					
HMDB05377	C50:2 TAG	Lipids and lipid-like molecules	2.06 (1.60, 2.65)	<0.0001	<0.001
HMDB07102	C34:1 DAG	Lipids and lipid-like molecules	2.03 (1.60, 2.58)	<0.0001	<0.001
HMDB05369	C52:2 TAG	Lipids and lipid-like molecules	1.99 (1.52, 2.61)	<0.0001	<0.001
HMDB07132	C34:3 DAG	Lipids and lipid-like molecules	1.83 (1.40, 2.41)	<0.0001	<0.001
HMDB07218	C36:2 DAG	Lipids and lipid-like molecules	1.81 (1.43, 2.29)	<0.0001	<0.001
HMDB07103	C34:2 DAG	Lipids and lipid-like molecules	1.80 (1.43, 2.28)	<0.0001	<0.001
HMDB05433	C50:3 TAG	Lipids and lipid-like molecules	1.72 (1.33, 2.22)	<0.001	<0.001
HMDB05384	C52:3 TAG	Lipids and lipid-like molecules	1.46 (1.14, 1.87)	0.003	0.01
HMDB05435	C50:4 TAG	Lipids and lipid-like molecules	1.39 (1.09, 1.76)	0.01	0.02
HMDB10471	C50:5 TAG	Lipids and lipid-like molecules	1.38 (1.10, 1.73)	0.01	0.01
HMDB07219	C36:3 DAG	Lipids and lipid-like molecules	1.36 (1.10, 1.70)	0.01	0.01
HMDB10497	C50:6 TAG	Lipids and lipid-like molecules	1.35 (1.09, 1.67)	0.01	0.01
HMDB10517	C52:7 TAG	Lipids and lipid-like molecules	1.17 (0.95, 1.45)	0.14	0.24
HMDB05436	C52:6 TAG	Lipids and lipid-like molecules	1.16 (0.93, 1.45)	0.20	0.28
HMDB10518	C54:8 TAG	Lipids and lipid-like molecules	0.99 (0.80, 1.22)	0.91	0.91
HMDB10498	C54:9 TAG	Lipids and lipid-like molecules	0.90 (0.73, 1.12)	0.34	0.39
HMDB05448	C56:9 TAG	Lipids and lipid-like molecules	0.89 (0.71, 1.10)	0.28	0.33
HMDB10513	C56:10 TAG	Lipids and lipid-like molecules	0.87 (0.71, 1.08)	0.20	0.28
HMDB11526	C22:6 LPE	Lipids and lipid-like molecules	0.87 (0.70, 1.07)	0.18	0.28
PCA (proportion)†					
First component (44.9%)			1.15 (1.08, 1.22)	<0.0001	
Second component (15.1%)			0.80 (0.71, 0.89)	<0.0001	
Third component (10.3%)			0.96 (0.86, 1.08)	0.49	

*Conditional logistic regression analysis was conducted to adjust for age at blood draw (continuous), race (White, non-White), cumulative average levels of BMI (<23.0, 23.0–24.9, 25.0–27.4, 27.5–29.9, ≥30.0 kg/m²), physical activity (<3.0, 3.0–8.9, 9.0–17.9, 18.0–26.9, ≥27.0 MET h/week), alcohol consumption (0, 0.1–4.9, 5.0–9.9, 10.0–14.9, ≥15.0 g/day), smoking status (never smoker; past smoker, <30 pack-years; past smoker, ≥30 pack-years; current smoker, <30 pack-years; current smoker, ≥30 pack-years), whole grains (quintiles), fruits (quintiles), vegetables (quintiles), polyunsaturated fat-to-saturated fat ratio (quintiles), fish (quintiles), red meat (quintiles), sugar-sweetened beverages (quintiles), total energy intake (quintiles), regular multivitamin use (yes, no), regular aspirin/nonsteroidal anti-inflammatory drug (NSAID) use (yes, no), family history of diabetes (yes, no), menopausal status (premenopausal, postmenopausal, unknown), and menopausal hormone therapy (never or past or current use). †PCA was performed for 30 coffee-related metabolites, and the top three components were linked to diabetes risk.

Coffee is known to impact lipid metabolism, but the biomarkers and specific pathways remain poorly understood. In our study, most of the measured metabolites were lipids, including CEs, DAGs, TAGs, phosphocholines, phosphatidylethanolamines, lysophosphatidylcholines, and lysophosphatidylethanolamines (LPEs), thus allowing us to identify specific lipid metabolites associated with long-term coffee intake. We observed that coffee intake was positively associated with three CEs (C20:4 CE, C18:1 CE, and C18:2 CE) and inversely associated with 21 lipid species, primarily DAGs and TAGs. In support of our findings, several observational studies have

reported an inverse association between coffee intake and TAG concentrations (22,23). Experimental evidence indicated that coffee polyphenols (e.g., chlorogenic acid) significantly lowered plasma free fatty acid, triglyceride, and total cholesterol concentrations, partly by inhibition of cholesterol synthesis and stimulation of fatty acid β -oxidation activity in the liver (24,25). This may explain our findings of the inverse associations of coffee intake with DAGs and TAGs.

High plasma TAG concentrations and low HDL cholesterol have been linked to increased diabetes risk, and dyslipidemia is likely caused by increased free fatty

acid flux secondary to insulin resistance (26). We showed that DAGs and TAGs negatively associated with coffee intake were associated with higher diabetes risk. Consistent with our findings, two prospective cohort studies found that adjustment for TAGs attenuated the association between coffee intake and diabetes risk (22,27). On the other hand, CEs positively associated with coffee showed an inverse association with diabetes risk. Similar findings have been reported in other studies (28). CEs are mainly located in plasma lipoproteins, particularly HDL, which delivers excess cholesterol from peripheral tissues to the

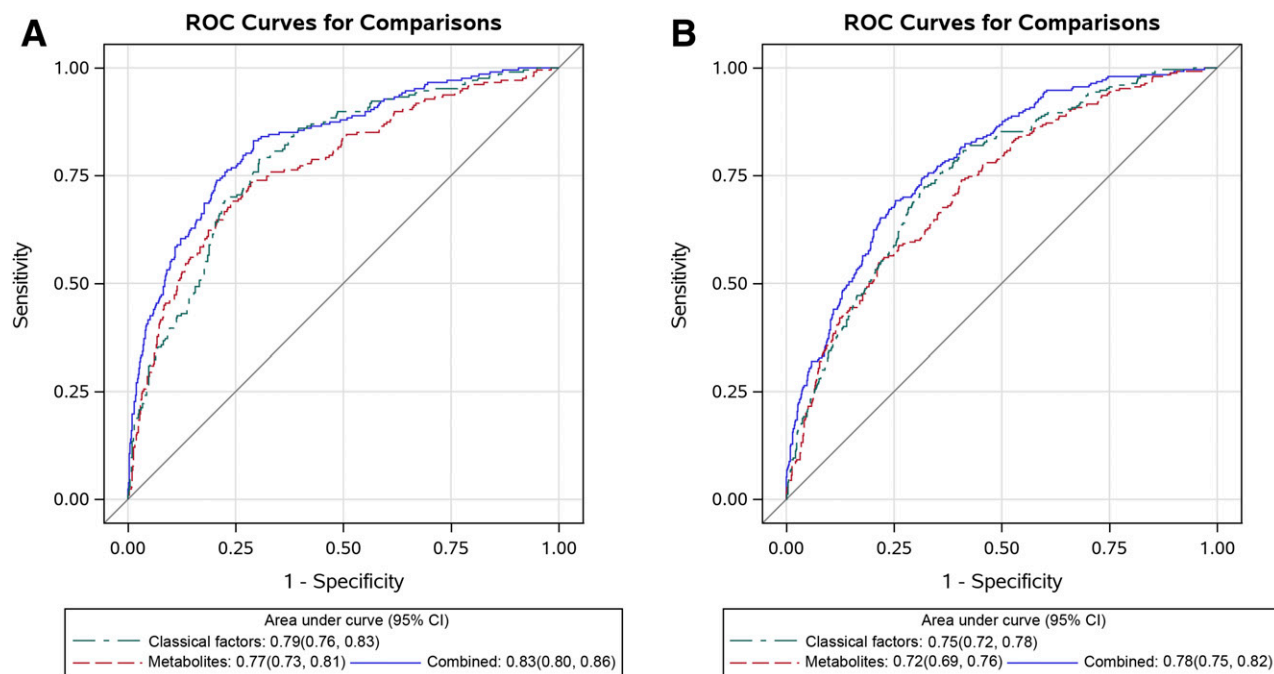


Figure 1—The C-statistics (95% CIs) of prediction models based on classical factors (including age, BMI, physical activity, smoking, and family history of diabetes), 15 (A) or 14 (B) plasma metabolites associated with type 2 diabetes risk, and the combination of classical factors and metabolites in the training set (A) and validation set (B). A: $P < 0.001$ for the combined model vs. classical factor model and $P = 0.35$ for the metabolite model vs. classical factor model. B: $P < 0.001$ for the combined model vs. classical factor model and $P = 0.23$ for the metabolite model vs. classical factor model.

liver for excretion (29). It is possible that coffee intake may reduce diabetes risk by elevating CE and HDL levels. Indeed, a clinical trial showed a significant increase in serum HDL cholesterol after consumption of 4–8 cups of coffee/day for 3 months (30). Alternatively, our observed associations between coffee and lipid metabolites may represent noncausal effects of insulin resistance, which is a direct cause of diabetes.

Consistent with previous studies (7,31), we found a strong association of total and caffeinated coffee intake with caffeine and three caffeine metabolites, including AAMU, 1,7-dimethyluric acid, and 7-methylxanthine. In a recent study in the Women's Health Initiative, these four metabolites were all associated with lower inflammatory potential of the diet, in line with the anti-inflammatory effect of coffee (18). The main route of caffeine metabolism is carried out by CYP1A2 and CYP2A6 in the liver, through N-3-demethylation to paraxanthine before catalyzation to AAMU, 1,7-dimethyluric acid, and 7-methylxanthine (32). Regarding the role of caffeine in diabetes development, while short-term human studies showed that caffeinated coffee intake might induce acute reduction in insulin

sensitivity (33,34), a meta-analysis of 28 prospective studies found that both caffeinated and decaffeinated coffee consumption was associated with a lower risk of diabetes (35). Our previous study also found that both caffeinated and decaffeinated coffee intake were associated with favorable profiles of various biomarkers in key metabolic and inflammatory pathways (3). These results are consistent with our current observation for a null association of caffeine and caffeine metabolites with diabetes risk, suggesting that components in coffee other than caffeine may drive the beneficial effect on diabetes.

Coffee is a rich source of polyphenols. We identified two coffee-related metabolites, cinnamoylglycine and 4-hydroxyhippuric acid, that are produced in microbial metabolism of polyphenols. Human gut contains a diversity of microbial populations. Specific bacterial enzymes can transform polyphenols into smaller metabolites through deglycosylation, dehydroxylation, and demethylation before these metabolites undergo further metabolism in systemic circulation (36). Cinnamoylglycine is the glycine conjugate of cinnamic acid, which is a gut microbial metabolite of polyphenols (37); 4-

hydroxyhippuric acid, another microbial product derived from polyphenol metabolism, is produced by hepatic conjugation of 4-hydroxybenzoic acid with glycine (38). In line with our study, these metabolites have previously been associated with higher consumption of polyphenol-rich foods including coffee (39). To our knowledge, no prior studies have assessed these polyphenol metabolites in relation to diabetes risk. In our nested case-control studies, only 4-hydroxyhippuric acid was available, and its association with diabetes was null. Further studies are needed to further examine the relationship between polyphenol metabolites and risk of diabetes.

Our study has several strengths, including the large sample size, repeated assessment of caffeinated and decaffeinated coffee intake prior to blood draw, collection of detailed covariate data for confounding control, and a multimetabolomics approach for analysis of a wide range of biochemical compounds. Moreover, we performed internal validation for the coffee metabolites and their associations with diabetes risk, using samples from the same source cohorts and the same metabolomic platforms in a single laboratory. Our study also has

several limitations. First, there may be measurement errors associated with dietary assessment by FFQs. However, the FFQ used in our cohorts has been validated for assessment of long-term coffee intake. Second, we did not have information on the coffee bean type, roast and preparation methods, or the amount of sugar/cream added into coffee, which may influence metabolic response to coffee consumption. For example, compared with unfiltered coffee, filtered coffee contained much less lipid content. During the follow-up period of our cohorts, most of the coffee consumed in the U.S. was filtered coffee. Thus, whether our findings are applicable to unfiltered coffee needs to be examined in future studies. Third, our study is observational and unable to establish causality. However, we performed rigorous multivariable adjustment to minimize the influence of residual confounding. Also, our observations are supported by biological evidence. Fourth, metabolomic profiling was conducted only once. However, our pilot study of repeated assessments over 1–2 years showed that levels of most metabolites were highly stable over time (intraclass correlation coefficient or Spearman correlation >0.65) (40). Fifth, the identified lipid species are interrelated with each other, making it difficult to disentangle their independent associations with coffee intake and diabetes risk. Finally, participants in our study were all women and predominantly Whites, limiting the generalizability of our findings. Further confirmation in men and other racial/ethnic groups is needed.

In summary, we identified a panel of 34 plasma metabolites associated with coffee intake. We also provide evidence that coffee may reduce diabetes risk by modulating lipid metabolism and that coffee-related metabolites may improve prediction of diabetes risk beyond classical risk factors. Future prospective and interventional studies are needed to confirm our findings.

Acknowledgments. The authors thank the participants and staff of the NHS and NHSII for their valuable contributions.

Funding. This work was supported by the American Cancer Society Mentored Research Scholar Grant (MRS-G-17-220-01-NEC [to M.S.]), by U.S. National Institutes of Health grants (UM1 CA186107 to M.J. Stampfer; R01 CA49449 to S.E. Hankinson; U01 CA176726 and R01 CA050385

to W.C.W. and A.H.E.; P01 CA087969 and R01 CA163451 to S.S. Tworoger; R01 AR057327 to E.W. Karlson; R01 NS045893 and R01 NS089619 to A. Ascherio; R01 CA087969 to R.M. Tamimi; K24 DK098311, R01 CA137178, R01 CA202704, and R01 CA176726 to A.T.C.; K99 CA215314 and R00 CA215314 to M.S.; and R01 DK112940 to F.B.H.), by the Department of Defense (W81XWH-12-1-0561), and by grants from National Natural Science Foundation of China (81973127 to D.H.) and Natural Science Foundation of Jiangsu Province (BK20190083 to D.H.). A.T.C. is a Stuart and Suzanne Steele MGH Research Scholar.

Grants to individuals who are not authors of this work contributed to the establishment of cohorts but not directly to the current analysis. The authors assume full responsibility for analyses and interpretation of these data.

Duality of Interest. No potential conflicts of interest relevant to this article were reported.

Author Contributions. D.H. performed statistical analysis and drafted the manuscript. O.A.Z., X.H., M.G.-F., X.J., J.L., L.L., A.H.E., C.B.C., A.T.C., Z.H., H.S., K.M.W., L.A.M., Q.S., F.B.H., W.C.W., and E.L.G. contributed to the acquisition of data, interpretation of the results, and revision of the manuscript. M.S. was responsible for study design. D.H. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

- Grosso G, Godos J, Galvano F, Giovannucci EL. Coffee, caffeine, and health outcomes: an umbrella review. *Annu Rev Nutr* 2017;37:131–156
- Millen BE, Abrams S, Adams-Campbell L, et al. The 2015 Dietary Guidelines Advisory Committee Scientific report: development and major conclusions. *Adv Nutr* 2016;7:438–444
- Hang D, Kværner AS, Ma W, et al. Coffee consumption and plasma biomarkers of metabolic and inflammatory pathways in US health professionals. *Am J Clin Nutr* 2019;109:635–647
- Ludwig IA, Clifford MN, Lean ME, Ashihara H, Crozier A. Coffee: biochemistry and potential impact on health. *Food Funct* 2014;5:1695–1717
- Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol* 2016;17:451–459
- Altmajer E, Kastenmüller G, Römisch-Margl W, et al. Variation in the human lipidome associated with coffee consumption as revealed by quantitative targeted metabolomics. *Mol Nutr Food Res* 2009;53:1357–1365
- Zheng Y, Yu B, Alexander D, Steffen LM, Boerwinkle E. Human metabolome associates with dietary intake habits among African Americans in the atherosclerosis risk in communities study. *Am J Epidemiol* 2014;179:1424–1433
- Playdon MC, Ziegler RG, Sampson JN, et al. Nutritional metabolomics and breast cancer risk in a prospective study. *Am J Clin Nutr* 2017;106:637–649
- Papandreou C, Hernández-Alonso P, Bulló M, et al. Plasma metabolites associated with coffee consumption: a metabolomic approach within the PREDIMED study. *Nutrients* 2019;11:1032
- Bao Y, Bertoia ML, Lenart EB, et al. Origin, methods, and evolution of the three Nurses' Health Studies. *Am J Public Health* 2016;106:1573–1581
- Hunter DJ, Hankinson SE, Hough H, et al. A prospective study of NAT2 acetylation genotype, cigarette smoking, and risk of breast cancer. *Carcinogenesis* 1997;18:2127–2132
- Salvini S, Hunter DJ, Sampson L, et al. Food-based validation of a dietary questionnaire: the effects of week-to-week variation in food consumption. *Int J Epidemiol* 1989;18:858–867
- Ding M, Satija A, Bhupathiraju SN, et al. Association of coffee consumption with total and cause-specific mortality in 3 large prospective cohorts. *Circulation* 2015;132:2305–2315
- Paynter NP, Balasubramanian R, Giulianini F, et al. Metabolic predictors of incident coronary heart disease in women. *Circulation* 2018;137:841–853
- Zeleznik OA, Eliassen AH, Kraft P, et al. A prospective analysis of circulating plasma metabolites associated with ovarian cancer risk. *Cancer Res* 2020;80:1357–1367
- Gönen M. *Analyzing Receiver Operating Characteristic Curves Using SAS*. Cary, NC, SAS Press, 2007
- Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128–138
- Tabung FK, Liang L, Huang T, et al. Identifying metabolomic profiles of inflammatory diets in postmenopausal women. *Clin Nutr* 2020;39:1478–1490
- Huxley R, Lee CM, Barzi F, et al. Coffee, decaffeinated coffee, and tea consumption in relation to incident type 2 diabetes mellitus: a systematic review with meta-analysis. *Arch Intern Med* 2009;169:2053–2063
- Bhupathiraju SN, Pan A, Malik VS, et al. Caffeinated and caffeine-free beverages and risk of type 2 diabetes. *Am J Clin Nutr* 2013;97:155–166
- van Dam RM, Willett WC, Manson JE, Hu FB. Coffee, caffeine, and risk of type 2 diabetes: a prospective cohort study in younger and middle-aged U.S. women. *Diabetes Care* 2006;29:398–403
- Jacobs S, Kröger J, Floegel A, 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
- Takami H, Nakamoto M, Uemura H, et al. Inverse correlation between coffee consumption and prevalence of metabolic syndrome: baseline survey of the Japan Multi-Institutional Collaborative Cohort (J-MICC) Study in Tokushima, Japan. *J Epidemiol* 2013;23:12–20
- Cho AS, Jeon SM, Kim MJ, et al. Chlorogenic acid exhibits anti-obesity property and improves lipid metabolism in high-fat diet-induced obese mice. *Food Chem Toxicol* 2010;48:937–943
- Wan CW, Wong CN, Pin WK, et al. Chlorogenic acid exhibits cholesterol lowering and fatty liver attenuating properties by up-regulating the gene expression of PPAR- α in hypercholesterolemic rats induced with a high-cholesterol diet. *Phytother Res* 2013;27:545–551
- Mooradian AD. Dyslipidemia in type 2 diabetes mellitus. *Nat Clin Pract Endocrinol Metab* 2009;5:150–159
- Rosengren A, Dotevall A, Wilhelmsen L, Thelle D, Johansson S. Coffee and incidence of diabetes in Swedish women: a prospective 18-year follow-up study. *J Intern Med* 2004;255:89–95

28. Razquin C, Toledo E, Clish CB, et al. Plasma lipidomic profiling and risk of type 2 diabetes in the PREDIMED trial. *Diabetes Care* 2018;41:2617–2624
29. Rousset X, Vaisman B, Amar M, Sethi AA, Remaley AT. Lecithin: cholesterol acyltransferase—from biochemistry to role in cardiovascular disease. *Curr Opin Endocrinol Diabetes Obes* 2009;16:163–171
30. Kempf K, Herder C, Erlund I, et al. Effects of coffee consumption on subclinical inflammation and other risk factors for type 2 diabetes: a clinical trial. *Am J Clin Nutr* 2010;91:950–957
31. Cornelis MC, Erlund I, Michelotti GA, Herder C, Westerhuis JA, Tuomilehto J. Metabolomic response to coffee consumption: application to a three-stage clinical trial. *J Intern Med* 2018;283:544–557
32. Thorn CF, Aklillu E, McDonagh EM, Klein TE, Altman RB. PharmGKB summary: caffeine pathway. *Pharmacogenet Genomics* 2012;22:389–395
33. Keijzers GB, De Galan BE, Tack CJ, Smits P. Caffeine can decrease insulin sensitivity in humans. *Diabetes Care* 2002;25:364–369
34. Battram DS, Arthur R, Weekes A, Graham TE. The glucose intolerance induced by caffeinated coffee ingestion is less pronounced than that due to alkaloid caffeine in men. *J Nutr* 2006;136:1276–1280
35. Ding M, Bhupathiraju SN, Chen M, van Dam RM, Hu FB. Caffeinated and decaffeinated coffee consumption and risk of type 2 diabetes: a systematic review and a dose-response meta-analysis. *Diabetes Care* 2014;37:569–586
36. Rowland I, Gibson G, Heinken A, et al. Gut microbiota functions: metabolism of nutrients and other food components. *Eur J Nutr* 2018;57:1–24
37. Wikoff WR, Anfora AT, Liu J, et al. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc Natl Acad Sci U S A* 2009;106:3698–3703
38. Rechner AR, Kuhnle G, Hu H, et al. The metabolism of dietary polyphenols and the relevance to circulating levels of conjugated metabolites. *Free Radic Res* 2002;36:1229–1241
39. Rothwell JA, Madrid-Gambin F, Garcia-Aloy M, et al. Biomarkers of intake for coffee, tea, and sweetened beverages. *Genes Nutr* 2018;13:15
40. Townsend MK, Clish CB, Kraft P, et al. Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. *Clin Chem* 2013;59:1657–1667