

# Walnut Consumption, Plasma Metabolomics, and Risk of Type 2 Diabetes and Cardiovascular Disease

Marta Guasch-Ferré,<sup>1,2</sup> Pablo Hernández-Alonso,<sup>3,4,5</sup> Jean-Philippe Drouin-Chartier,<sup>1,6</sup> Miguel Ruiz-Canela,<sup>5,7</sup> Cristina Razquin,<sup>5,7</sup> Estefanía Toledo,<sup>5,7</sup> Jun Li,<sup>1,8</sup> Courtney Dennis,<sup>9</sup> Clemens Wittenbecher,<sup>1,10,11</sup> Dolores Corella,<sup>5,12</sup> Ramon Estruch,<sup>5,13</sup> Montserrat Fitó,<sup>5,14</sup> Emilio Ros,<sup>5,15</sup> Nancy Babio,<sup>3,4,5</sup> Shilpa N Bhupathiraju,<sup>1,2</sup> Clary B Clish,<sup>9</sup> Liming Liang,<sup>8,16</sup> Miguel A Martínez-González,<sup>1,5,7</sup> Frank B Hu,<sup>1,2,8</sup> and Jordi Salas-Salvadó<sup>3,4,5</sup>

<sup>1</sup>Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA; <sup>2</sup>Channing Division for Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; <sup>3</sup>Universitat Rovira i Virgili, Departament de Bioquímica i Biotecnologia, Unitat de Nutrició Humana, Hospital Universitari San Joan de Reus, Reus, Spain; <sup>4</sup>Institut d'Investigació Pere Virgili, Reus, Spain; <sup>5</sup>Consorcio CIBER, M.P. Fisiopatología de la Obesidad y Nutrición, Instituto de Salud Carlos III, Madrid, Spain; <sup>6</sup>Centre Nutrition, Santé et Société, Institut sur la Nutrition et les Aliments Fonctionnels, Faculté de Pharmacie, Université Laval, Québec, Canada; <sup>7</sup>Department of Preventive Medicine and Public Health, Navarra Health Research Institute (IDISNA), University of Navarra, Pamplona, Spain; <sup>8</sup>Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA; <sup>9</sup>The Broad Institute of Harvard and MIT, Boston, MA, USA; <sup>10</sup>Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany; <sup>11</sup>German Center for Diabetes Research, Neuherberg, Germany; <sup>12</sup>Department of Preventive Medicine, University of Valencia, Valencia, Spain; <sup>13</sup>Department of Internal Medicine, Institut d'Investigacions Biomèdiques August Pi Sunyer, Hospital Clinic, University of Barcelona, Barcelona, Spain; <sup>14</sup>Cardiovascular and Nutrition Research Group, Institut de Recerca Hospital del Mar, Barcelona, Spain; <sup>15</sup>Lipid Clinic, Department of Endocrinology and Nutrition, Agust Pi i Sunyer Biomedical Research Institute (IDIBAPS), Hospital Clinic, University of Barcelona, Barcelona, Spain; and <sup>16</sup>Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

## ABSTRACT

**Background:** Walnut consumption is associated with lower risk of type 2 diabetes (T2D) and cardiovascular disease (CVD). However, it is unknown whether plasma metabolites related to walnut consumption are also associated with lower risk of cardiometabolic diseases.

**Objectives:** The study aimed to identify plasma metabolites associated with walnut consumption and evaluate the prospective associations between the identified profile and risk of T2D and CVD.

**Methods:** The discovery population included 1833 participants at high cardiovascular risk from the PREvención con Dleta MEDiterránea (PREDIMED) study with available metabolomics data at baseline. The study population included 57% women (baseline mean BMI (in kg/m<sup>2</sup>): 29.9; mean age: 67 y). A total of 1522 participants also had available metabolomics data at year 1 and were used as the internal validation population. Plasma metabolomics analyses were performed using LC-MS. Cross-sectional associations between 385 known metabolites and walnut consumption were assessed using elastic net continuous regression analysis. A 10-cross-validation (CV) procedure was used, and Pearson correlation coefficients were assessed between metabolite weighted models and self-reported walnut consumption in each pair of training-validation data sets within the discovery population. We further estimated the prospective associations between the identified metabolite profile and incident T2D and CVD using multivariable Cox regression models.

**Results:** A total of 19 metabolites were significantly associated with walnut consumption, including lipids, purines, acylcarnitines, and amino acids. Ten-CV Pearson correlation coefficients between self-reported walnut consumption and the plasma metabolite profile were 0.16 (95% CI: 0.11, 0.20) in the discovery population and 0.15 (95% CI: 0.10, 0.20) in the validation population. The metabolite profile was inversely associated with T2D incidence (HR per 1 SD: 0.83; 95% CI: 0.71, 0.97; *P* = 0.02). For CVD incidence, the HR per 1-SD was 0.71 (95% CI: 0.60, 0.85; *P* < 0.001).

**Conclusions:** A metabolite profile including 19 metabolites was associated with walnut consumption and with a lower risk of incident T2D and CVD in a Mediterranean population at high cardiovascular risk. *J Nutr* 2021;151:303–311.

**Keywords:** walnuts, metabolomics, type 2 diabetes, cardiovascular disease, walnut consumption

## Introduction

Walnuts are one of the most popular types of nuts consumed worldwide. Walnuts are a good source of PUFAs and are rich in fiber, nonsodium minerals (i.e., potassium, calcium, and magnesium), vitamins, phytosterols, and polyphenols (1). Compared to other types of nuts, walnuts contain higher amounts of the essential fatty acid  $\alpha$ -linolenic acid (ALA; 18:3n-3) (11.6% of total fatty acid composition for walnuts compared to <0.7% in other nuts) and linoleic acid (LA; 18:2n-6) (2). Once absorbed, LA, the main vegetable n-3 PUFA, is only modestly converted to its longer chain counterparts EPA and DHA, but it appears to possess anti-inflammatory and antiatherogenic properties on its own (3). The unique fatty acid composition together with their richness in phytosterols are believed to underlie the consistent cholesterol-lowering effect of regular walnut consumption (4). In addition, accumulating evidence from prospective studies suggests that higher walnut consumption is associated with lower risk of various chronic diseases, including type 2 diabetes (T2D) and cardiovascular disease (CVD) (5, 6).

Although current evidence has shown a wide range of health benefits from walnut consumption, the biological mechanisms underlying these salutary effects are not fully understood. Nutritional metabolomics is a rapidly evolving approach to obtain deeper insights into diet-disease association that holds great promise in improving our understanding of the biological

effects of nutritional factors and may also help identify potential novel biomarkers of dietary intake and/or disease risk prediction (7).

A few previous studies, including acute feeding studies, clinical trials, and observational studies, have evaluated how the consumption of walnuts influences plasma and urinary metabolites (8). Whereas some metabolites, including ALA-derived oxylipins, (2) urolithins (9), and 5-hydroxyindole-3-acetic acid, have been identified to be associated with walnut consumption, findings for other metabolites have been inconsistent (8). Furthermore, no previous studies have investigated whether plasma metabolites correlated with walnut intake are also associated with a lower risk of cardiometabolic diseases.

In the current study, we used an agnostic machine learning approach to identify plasma metabolites associated with walnut consumption using data from the PREVENCIÓN con Dieta MEDiterránea (PREDIMED) study. We then assessed whether the identified metabolite profile was associated with T2D and CVD incidence risk independently of known risk factors.

## Methods

### Study population

#### Discovery population.

The study cohort was derived from the PREDIMED study, a multicenter randomized feeding trial conducted in Spain from 2003 to 2010, which examined the effects of the traditional Mediterranean diet (MedDiet) in the primary prevention of CVD in a population at high cardiovascular risk. A detailed description of the PREDIMED trial can be found elsewhere (10, 11). The protocol was approved by the Institutional Review Boards at all PREDIMED study locations, and all participants provided written informed consent.

Two nested case-cohort studies were designed for metabolomics profiling within the PREDIMED trial—1 for CVD, which was the primary outcome of the trial, and 1 for T2D, which was a secondary outcome. The PREDIMED-CVD study consisted of 229 incident CVD cases free of CVD at baseline and 788 subcohort participants (overlapping  $n = 37$ ) (12, 13), and the PREDIMED-T2D study consisted of 251 incident T2D cases and 641 subcohort participants (overlapping  $n = 53$ ) without T2D at baseline (14, 15). Participants with oral glucose tolerance data at baseline ( $n = 130$ ) were also included in the PREDIMED-T2D project (Supplemental Figure 1).

For the current study, participants with available baseline metabolomics data from the 2 case-cohort studies and with a completed validated semiquantitative 137-item FFQ were selected to develop the metabolite profile model ( $n = 1882$ ). Participants with missing FFQ data at baseline ( $n = 11$ ), those who had daily energy intake <500 or >3500 kcal/d for women and <800 or >4000 kcal/d for men ( $n = 34$ ), and those participants with  $\geq 20\%$  missing values in metabolites ( $n = 4$ ) were excluded. A total of 1833 participants were included in the discovery population of the current analysis. From the 1833 participants at baseline, 571 were randomly assigned to the control diet, 633 were randomly assigned to the MedDiet supplemented with extra virgin olive oil, and 639 participants were randomly assigned to the MedDiet supplemented with nuts.

#### Validation population.

Internal validation was conducted using data from year 1 in the PREDIMED study. Of the 1833 participants included in the discovery population, 1522 had repeated measurements of diet and metabolomics at year 1 of intervention, and they were used as the validation set (Supplemental Figure 1).

#### Dietary assessment

At baseline and yearly thereafter, trained dietitians completed a validated 137-item semiquantitative FFQ in face-to-face interviews

---

The PREDIMED study was funded by NIH grants R01 HL118264 and R01 DK102896 and by the Spanish Ministry of Health (Instituto de Salud Carlos III, The PREDIMED Network grant RD 06/0045, 2006–2013, coordinated by MAM-G; and a previous network grant RTIC-G03/140, 2003–2005, coordinated R. Estruch). Additional grants were received from the Ministerio de Economía y Competitividad-Fondo Europeo de Desarrollo Regional (Proyectos CNIC-06/2007, CIBER 06/03, PI06-1326, PI07-0954, PI11/02505, SAF2009-12304, and AGL2010-22319-C03-03) and the Generalitat Valenciana (ACOMP2010-181, AP-111/10, AP-042/11, ACOM2011/145, ACOMP/2012/190, ACOMP/2013/159, ACOMP/213/165, and PROMETEO17/2017). MG-F is supported by American Diabetes Association grant 1-18-PMF-029. PH-A is supported by a postdoctoral fellowship (Juan de la Cierva-Formación, FJCI-2017-32205). CW was supported by an individual fellowship from the German Research Foundation (DFG). JS-S gratefully acknowledges the financial support by ICREA under the ICREA Academia program.

Author disclosures: J-PD-C received speaker and consulting honoraria from the Dairy Farmers of Canada in 2016 and 2018, outside the submitted work. ER reports grants, personal fees, nonfinancial support, and other from the California Walnut Commission. JS-S reports serving on the board of the International Nut and Dried Fruit Council and receiving grant support from this entity through his institution. He also reports serving on the Executive Committee of the Instituto Danone, Spain. He has also received research funding (tree nuts and olive oil for the PREDIMED and/or PREDIMED-Plus trial) from the California Walnut Commission; the Almond Board of California; Patrimonio Comunal Olivarero, Spain; La Morella Nuts, Spain; and Borges S.A., Spain. He reports receiving consulting fees or travel expenses outside of the submitted work from Danone, Spain; Eroski Foundation, Spain; the International Nut and Dried Fruit Council, Spain; and the Australian Nut Industry Council, Australia. FBH received research support from the California Walnut Commission. All other authors report no conflicts of interest. The funding sources had no role in the design, collection, analysis, or interpretation of the data or in the decision to submit the manuscript for publication.

Supplemental Figure 1 and Supplemental Tables 1–3 are available from the “Supplementary data” link in the online posting of the article and from the same link in the online table of contents at <https://academic.oup.com/ajcn/>.

MG-F, PH-A, FBH, and JS-S contributed equally to this work.

Address correspondence to MG-F (e-mail: [mguasch@hsph.harvard.edu](mailto:mguasch@hsph.harvard.edu)) or PH-A (e-mail: [pablo.hernandez@alumni.urv.cat](mailto:pablo.hernandez@alumni.urv.cat)).

Abbreviations used: ALA,  $\alpha$ -linolenic acid; CV, cross-validation; CVD, cardiovascular disease; LA, linoleic acid; MAG, monoacylglycerol; MedDiet, Mediterranean diet; MSE, mean squared error; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PREDIMED, PREVENCIÓN con Dieta MEDiterránea; PS, phosphatidylserine; T2D, type 2 diabetes.

with the participants. Energy and nutrient intake were estimated using Spanish food composition tables (16, 17). Information on self-reported walnut consumption was derived from the FFQ. The questionnaire includes 1 item regarding the specific consumption of walnuts. The dietitians asked the participants how often they consumed walnuts, ranging from never to between 1 and 3 times per month, and how many times per week (1, 2–4, or 5–6 times) or times per day (1, 2–3, 4–6, or >6 times) they consumed walnuts.

### Assessment of risk factors and covariates

Medical conditions, family history of disease, and risk factors were collected through a questionnaire during the first screening visit. At baseline and during annual visits, trained personnel measured participants' body weight, height, waist circumference, and blood pressure according to the study protocol. Physical activity was assessed using the validated Spanish version of the Minnesota Leisure-Time Physical Activity questionnaire (18). Participants were considered to have hypercholesterolemia or hypertension when they had previously been diagnosed and/or they were being treated with cholesterol-lowering or antihypertensive agents, respectively.

### Metabolite profiling

The plasma metabolomics profiling was performed at the Broad Institute of Harvard University and Massachusetts Institute of Technology using high-throughput LC-MS/MS techniques. After quality filtering and standardization, 399 named metabolites were qualified for primary analyses. From the 399 originally annotated metabolites, 11 metabolites were removed due to high number of missing values (i.e., >20%) and 3 metabolites were considered an internal standard, thus leaving a total of 385 metabolites in the final analysis.

All analyses used overnight fasting (fasting for  $\geq 8$  h) plasma EDTA samples collected at baseline and at year 1. Samples were processed at each recruiting center no later than 2 h after collection and stored in  $-80^{\circ}\text{C}$  freezers. Pairs of samples (baseline and first-year visit) from cases and subcohort participants were randomly distributed before being shipped to the Broad Institute for metabolomics assays.

LC-MS/MS was used to quantitatively profile polar metabolites and lipids of the plasma samples. Details of the LC-MS/MS platform can be found elsewhere (19–21). Briefly, amino acids and other polar metabolites were profiled with a Nexera X2 U-HPLC (Shimadzu) coupled to a Q-Exactive mass spectrometer (ThermoFisher Scientific). Metabolites were extracted from plasma (10  $\mu\text{L}$ ) using 90  $\mu\text{L}$  of 74.9:24.9:0.2 (vol:vol:vol) of acetonitrile/methanol/formic acid containing stable isotope-labeled internal standards [valine-d8 (Sigma-Aldrich) and phenylalanine-d8 (Cambridge Isotope Laboratories)]. The samples were centrifuged (9000  $\times g$ ; 10 min;  $4^{\circ}\text{C}$ ), and the supernatants were injected directly onto a 150  $\times$  2-mm, 3- $\mu\text{m}$  Atlantis HILIC column (Waters). The column was eluted isocratically at a flow rate of 250  $\mu\text{L}/\text{min}$  with 5% mobile phase A (10 mmol ammonium formate/L and 0.1% formic acid in water) for 0.5 min followed by a linear gradient to 40% mobile phase B (acetonitrile with 0.1% formic acid) over 10 min. MS analyses were carried out using electrospray ionization in the positive-ion mode, and full-scan spectra were acquired over 70–800  $m/z$ . Lipids were profiled using a Nexera X2 U-HPLC (Shimadzu) coupled to an Exactive Plus orbitrap MS (Thermo Fisher Scientific). Lipids were extracted from plasma (10  $\mu\text{L}$ ) using 190  $\mu\text{L}$  of isopropanol containing 1,2-didodecanoyl-sn-glycero-3-phosphocholine (Avanti Polar Lipids) as an internal standard. Lipid extracts (2  $\mu\text{L}$ ) were injected onto a 100  $\times$  2.1-mm, 1.7- $\mu\text{m}$  ACQUITY BEH C8 column (Waters). The column was eluted isocratically with 80% mobile-phase A (95:5:0.1 vol:vol:vol 10 mM ammonium acetate/methanol/formic acid) for 1 min followed by a linear gradient to 80% mobile-phase B (99.9:0.1 vol:vol methanol/formic acid) over 2 min, a linear gradient to 100% mobile-phase B over 7 min, and then 3 min at 100% mobile-phase B. MS analyses were carried out using electrospray ionization in the positive-ion mode using full-scan analysis over 200–1100  $m/z$ . Raw data were processed using Trace Finder version 3.1 and 3.3 (Thermo Fisher Scientific) and Progenesis QI (Nonlinear Dynamics). Polar metabolite identities were confirmed using authentic reference standards, and lipids

were identified by head group and total acyl carbon number and total acyl double bond content. To enable assessment of data quality and to facilitate data standardization across the analytical queue and sample batches, pairs of pooled plasma reference samples were analyzed at intervals of 20 study samples. One sample from each pair of pooled references served as a passive quality control sample to evaluate the analytical reproducibility for measurement of each metabolite, whereas the other pooled sample was used to standardize using a “nearest neighbor” approach. Standardized values were calculated using the ratio of the value in each sample over the nearest pooled plasma reference multiplied by the median value measured across the pooled references.

### Statistical analysis

To identify the metabolite profile associated with walnut consumption, we used walnut consumption and plasma metabolomics data from PREDIMED at baseline as the training set (discovery population). A combined data set from PREDIMED-CVD and -T2D nested case-control studies was used to increase sample size and ensure sufficient statistical power and model precision. Data from PREDIMED year 1 were used as the testing sets (validation cohorts) (Supplemental Figure 1).

Baseline characteristics of study participants were described as means and SDs for quantitative variables and as percentages for categorical variables. Missing values of individual metabolites were imputed in those metabolites with <20% of missing values using the random forest imputation approach (“missForest” function from the “missForest” R package) as recommended in metabolomics studies (22–24). Importantly, different alternatives to this approach were found to generate consistent results, as previously reported by our research consortium (25). Missing values correspond to those determinations that were below the limit of detection. To conduct the multivariate analysis, metabolomic data were first centered and scaled using the SD as the scaling factor (i.e., autoscaling) (26). Due to the high dimensionality and collinear nature of the data, Gaussian (i.e., continuous) linear regression with elastic net penalty (implemented in the “glmnet” R package) was used to build the model for walnut consumption. In the discovery population (PREDIMED baseline), we performed a 10 cross-validation (CV) approach—splitting the sample into training (90% of the sample) and validation sets (10% of the sample)—and then within the training sets we performed a further 10-fold CV to find the optimal value of the tuning parameter ( $\lambda$ ) that yielded the minimum mean squared error (MSE). The values  $\text{minMSE}$  and  $\text{minMSE} + 1 \text{ SE}$  were calculated using argument  $s = \text{“lambda.min”}$  or  $s = \text{“lambda.1se”}$  in the  $\text{cv.glmnet}$  function (“glmnet” R package). To report the coefficients from each CV iteration, we evaluated the  $\lambda$  selection in the elastic net logistic regression. We selected  $s = \text{lambda.min}$  because it gives the minimum mean CV error and  $s = \text{“lambda.1se”}$  was not deriving a valid model. In addition to the  $\lambda$  value, we evaluated the  $\alpha$  parameter from 0 (i.e., Ridge regression) to 1 (i.e., Lasso regression) in 0.1 increments to test the best scenario for our data. In this case,  $\alpha = 1$  was the model with best predicting accuracy in the validation sets. Weighted models were constructed for each training-validation data set pair (90% training and 10% validation) using the coefficients for the metabolites obtained from each elastic net regression in the training set.

Pearson correlation coefficients were derived between self-reported walnut consumption and the metabolite profile model in both the discovery and validation populations. For reproducibility purposes, regression coefficients are reported using 10 iterations of the 10-CV elastic regression approach in the whole data set. Metabolites that were selected in 9 out of the 10 iterations in the 10-CV elastic regression are also reported but not included in the main metabolite profile model. These analyses are based on consistency among CV runs, and therefore any  $P$ -value is derived.

We used weighted Cox regressions with Barlow weights and robust variance estimator to assess the associations between the metabolite profile associated with walnut consumption at baseline and year 1 with incident T2D risk (245 incident events from baseline and 161 events from year 1) and CVD risk (222 incident events from baseline and 151

incident events from year 1) within the T2D and CVD nested case-cohort studies, respectively. Multivariable model 1 (the basic model) was adjusted for age, sex, and propensity scores (11) and stratified by intervention group and recruitment center. Model 2 (sociodemographic model) was further adjusted for BMI, smoking status (never, former, or current smoker), alcohol intake (grams per day) and squared alcohol intake, education level (primary, secondary, academic), physical activity (metabolic-equivalent minutes per day), family history of coronary heart disease (yes/no), baseline dyslipidemia or lipid-lowering medication use (yes/no), baseline hypertension or antihypertensive use (yes/no), and T2D prevalence (only in CVD analyses). Model 3 (diet model) was additionally adjusted for total energy and intakes of vegetables, fruits, cereals, red and processed meat, fish, olive oil, eggs, legumes, and dairy in quintiles. Dietary variables were included on the basis of their association with the outcomes (CVD and T2D) and/or their correlation with the exposure (walnut intake). The final model [model 4 (+ self-reported walnut consumption)] included covariates from model 3 plus the consumption of walnuts from which the metabolite set was derived.

In sensitivity analysis, to test whether the associations were consistent if only metabolites positively associated with walnut intake are included in the models, we performed elastic net regression allowing only positive associations between metabolites and walnut consumption and calculated the 10-CV Pearson correlation coefficient in the discovery population using the metabolite profile model obtained.  $P < 0.05$  was considered statistically significant. All analyses were performed using R version 3.4.2 statistical software (R Foundation for Statistical Computing).

## Results

### Characteristics of the study participants

Baseline characteristics of the study participants are presented in [Table 1](#) by total population and by extremes of walnut consumption using tertiles. The mean  $\pm$  SD consumption of walnuts at baseline was  $6 \pm 9$  g/d and the respective consumption at 1 y was  $10 \pm 12$  g/d; walnut consumption increased due to the nature of the PREDIMED intervention. Participants with a higher consumption of walnuts at baseline had a lower BMI, were less likely to be women and to smoke, and had a higher intake of total energy ([Table 1](#)).

### Identification of walnut-related metabolites

[Figure 1](#) shows the metabolites selected 10 times in 10-CV validations of the elastic net regressions sorted by coefficient values. [Table 2](#) summarizes the number of metabolites and the Pearson correlation coefficients between the metabolite profile model and walnut consumption values in the discovery cohort and in the internal validation population (PREDIMED year 1). A total of 19 metabolites were significantly associated with walnut consumption. Of these, 9 metabolites had positive coefficients, whereas 10 had negative coefficients. The means and SDs of the metabolites' regression coefficients selected 9 and 10 times in the 10-CV elastic net regressions using lambda.min are shown in [Supplemental Table 1](#). In addition to the 19 metabolites that were selected 10 times in the 10-CV elastic net regressions, 15 additional metabolites, 9 with positive coefficients and 6 with negative coefficients, were selected a total of 9 times in the 10-CV elastic net regressions ([Supplemental Table 2](#)). [Table 3](#) shows the individual Pearson correlation coefficients between each consistently selected metabolite and walnut consumption. The direction of the associations between individual correlations is consistent with the coefficients of the metabolite profile model ([Figure 1](#)).

The 10:2 carnitine, 36:4 phosphatidylcholine (PC), biliverdin, guanine, 4-hydroxy-3-methylacetophenone, and

22:1 monoacylglycerol (MAG) were the metabolites with the strongest positive coefficient values, whereas isocitrate, 40:6 phosphatidylserine (PS), cyclohexylamine, succinate, and 36:5 phosphatidylethanolamine plasmalogen were the metabolites with strongest negative coefficient value ([Figure 1](#)). The Pearson correlations between walnut consumption and the metabolite profile were 0.16 (95% CI: 0.11, 0.20) in the discovery population (baseline) and 0.15 (95% CI: 0.10, 0.20) in the validation population (at year 1).

In sensitivity analysis, we calculated the Pearson correlation coefficients between self-reported walnut consumption and the metabolite profile obtained in the discovery population only allowing metabolites with positive coefficients, and the results were consistent with those of the primary analyses in both the discovery (0.13; 95% CI: 0.08, 0.19) and validation populations (0.14; 95% CI: 0.10, 0.19) ([Table 2](#)).

### Association of walnut-related metabolites with T2D and CVD risk

[Supplemental Table 3](#) depicts the characteristics of the study population included in the Cox models by T2D or CVD incident case-control status. T2D and CVD incident cases were more likely to be men and current smokers, and they had a higher BMI.

After adjusting for lifestyle and dietary risk factors, the HR and 95% CI for T2D per SD increment in the metabolite profile model of walnut consumption was 0.86 (95% CI: 0.74, 1.00;  $P = 0.05$ ) in the discovery population (cases = 245) (baseline data) and 0.82 (95% CI: 0.64, 1.05;  $P = 0.13$ ) in the internal validation population (cases = 161) (year 1 data) ([Table 4](#), model 3). When the models were further adjusted for self-reported walnut consumption (model 4), the results remained consistent.

The metabolite profile of walnut consumption was inversely associated with CVD risk in the discovery and in the internal validation populations after adjusting for potential confounders and self-reported walnut consumption ([Table 4](#)). The HR (95% CI) for CVD of the multivariable model adjusted for lifestyle and dietary factors was 0.66 (95% CI: 0.56, 0.77;  $P < 0.001$ ) using baseline (cases = 222) and 0.75 (95% CI: 0.60, 0.93;  $P = 0.01$ ) in the internal validation discovery population (cases = 159). The results remained significant when the models were adjusted for self-reported walnut consumption.

## Discussion

Using an agnostic machine-learning approach, we identified 19 plasma metabolites, including lipids, purines, acylcarnitines, and amino acids, that were associated with walnut consumption using data from the PREDIMED study. The identified walnut-related metabolite profile was inversely associated with T2D and CVD risk after adjusting for potential confounders and self-reported walnut consumption. Although the correlations between walnut consumption and the metabolite profile were weak, these findings may help provide new insights into the potential biological mechanisms underlying associations between walnuts and cardiometabolic health and illustrate the potential of metabolomics profiling to better understand mechanisms underlying diet-disease relationships. To our knowledge, this is the first study to specifically examine the association between plasma metabolite profile of walnut consumption and the risk of cardiometabolic diseases.



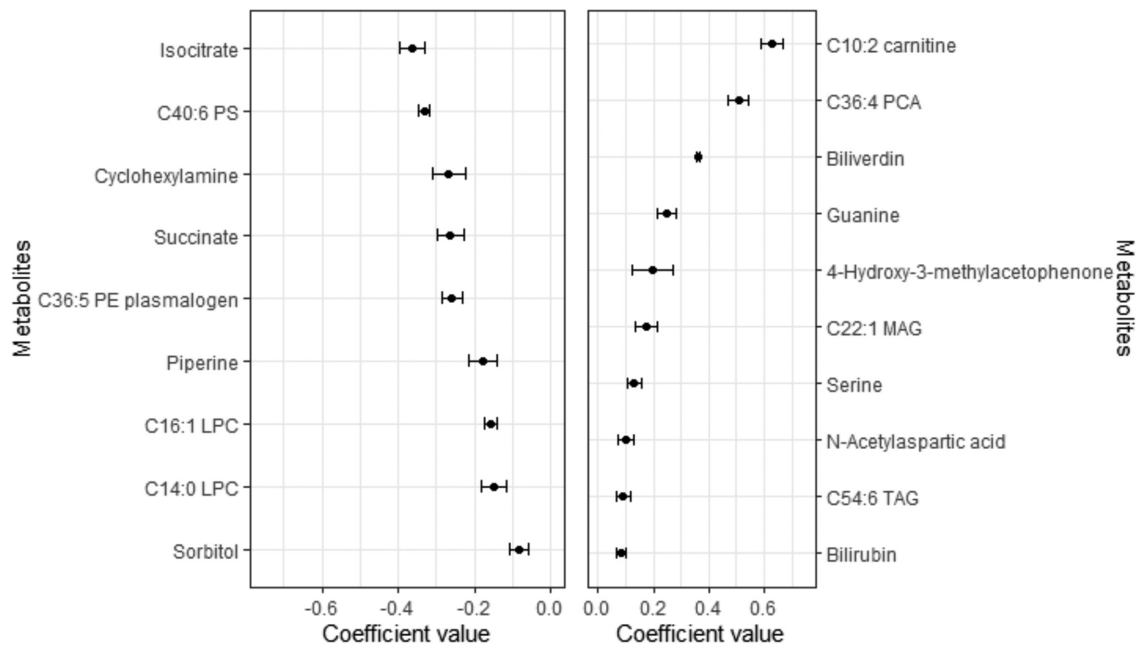
**TABLE 1** Characteristics of participants in the PREDIMED trial at baseline and year 1<sup>1</sup>

Characteristic	All participants	Participants with low walnut intake (tertile 1)	Participants with high walnut intake (tertile 3)
PREDIMED—baseline (discovery cohort)			
Participants, <i>n</i>	1833	691	467
Age, y	67 ± 6	67 ± 6	67 ± 6
Female sex, <i>n</i> (%)	1055 (57.6)	420 (60.8)	247 (52.9)
Prevalent type 2 diabetes, <i>n</i> (%)	492 (26.8)	209 (30.2)	110 (23.6)
BMI, kg/m <sup>2</sup>	29.9 ± 3.5	30.4 ± 3.7	29.2 ± 3.3
Physical activity, Mets-min/wk	254 ± 255	215 ± 226	278 ± 243
Current smoker, <i>n</i> (%)	287 (15.7)	106 (15.3)	67 (14.3)
History of hypercholesterolemia, <i>n</i> (%)	1408 (76.8)	513 (74.2)	377 (80.7)
History of hypertension, <i>n</i> (%)	1599 (87.2)	602 (87.1)	400 (85.7)
Family history of coronary heart disease, <i>n</i> (%)	451 (24.6)	172 (24.9)	106 (22.7)
Dietary intake			
Energy intake, kcal/d	2283 ± 544	2150 ± 538	2446 ± 541
Walnuts intake, g/d	6 ± 9	0 ± 0	19 ± 8
Total nuts intake, g/d	11 ± 13	2 ± 7	28 ± 14
Total dairy intake, g/d	375 ± 221	364 ± 21	379 ± 230
Vegetable intake, g/d	332 ± 150	312 ± 35	359 ± 172
Fruit intake, g/d	361 ± 197	340 ± 198	401 ± 198
Legume intake, g/d	20 ± 13	20 ± 14	21 ± 11
Grain intake, g/d	231 ± 101	220 ± 104	236 ± 100
Meat intake, g/d	134 ± 56	132 ± 57	132 ± 57
Fish intake, g/d	101 ± 53	92 ± 47	105 ± 49
Egg intake, g/d	20 ± 11	19 ± 11	19 ± 10
Alcohol intake, g/d	9 ± 15	8 ± 15	11 ± 16
PREDIMED—year 1 (internal replication cohort)			
Participants, <i>n</i>	1522	531	347
Age, y	68 ± 6	68 ± 6	67 ± 6
Female sex, <i>n</i> (%)	875 (57.5)	320 (60.3)	190 (54.8)
BMI, kg/m <sup>2</sup>	29.8 ± 3.7	30.3 ± 3.8	29.3 ± 3.6
Prevalent type 2 diabetes, <i>n</i> (%)	463 (30.4)	190 (35.8)	104 (30.0)
Physical activity, Mets-min/wk	247 ± 255	230 ± 244	224 ± 211
Current smoker, <i>n</i> (%)	215 (14.3)	81 (15.4)	52 (15.0)
History of hypercholesterolemia, <i>n</i> (%)	1153 (75.8)	383 (72.1)	266 (76.7)
History of hypertension, <i>n</i> (%)	1318 (86.6)	457 (86.1)	294 (84.7)
Dietary intake			
Energy intake, kcal/d	2285 ± 544	2077 ± 510	2574 ± 536
Walnuts intake, g/d	11 ± 12	0.5 ± 0.9	30 ± 10
Total nuts intake, g/d	21 ± 23	3 ± 5	55 ± 22
Total dairy intake, g/d	366 ± 214	364 ± 220	369 ± 211
Vegetable intake, g/d	345 ± 137	329 ± 140	381 ± 145
Fruit intake, g/d	389 ± 195	368 ± 196	422 ± 223
Legume intake, g/d	22 ± 11	22 ± 11	23 ± 10
Grain intake, g/d	225 ± 93	215 ± 96	232 ± 91
Meat intake, g/d	124 ± 51	122 ± 51	127 ± 56
Fish intake, g/d	106 ± 45	98 ± 45	116 ± 46
Egg intake, g/d	20 ± 11	20 ± 12	21 ± 9
Alcohol intake, g/d	8 ± 13	7 ± 12	9 ± 13

<sup>1</sup>Values are means ± SDs for continuous variables or number and percentages for categorical variables. Mets, metabolic-equivalent hours; PREDIMED, PREvención con Dieta MEDiterránea.

Some previous observational and intervention studies evaluated how the metabolome is influenced by walnut consumption. Traditionally, ALA has been considered the best blood biomarker of walnut consumption, mainly because one of the key differences in the fatty acid profile of walnuts compared with other nuts and other foods is the high amount of ALA in walnuts (2). Precisely, ALA was the biomarker used in the PREDIMED study to demonstrate compliance with the MedDiet supplemented with nuts (which included 15 g of

walnuts per day) (11). Acute feeding studies and clinical trials have identified high ALA in blood (including plasma and erythrocytes) after walnut consumption or when mixed nuts included walnuts; other studies have also identified LA after the consumption of walnuts, which is the major PUFA present in walnuts (8). McKay et al. (27) showed that compared to baseline concentrations, both LA and ALA of erythrocytes were higher after a 6-wk intervention with 42 g/d of walnut intake in 21 healthy men and women. However, other food sources



**FIGURE 1** Coefficients for the metabolites selected 10 times in the 10 cross-validations of the continuous elastic regression for walnut consumption in the PREDIMED study. Values are means  $\pm$  SDs for the set of metabolites consistently selected (i.e. 10 times) after 10 iterations of the elastic continuous regression procedure with 10-fold cross-validation (using lambda.min) employing the whole data set of subjects ( $n = 1833$ ). Metabolites with negative coefficients ( $n = 9$ ) are plotted on the left, whereas those with positive coefficients are shown on the right ( $n = 10$ ). LPC, lysophosphatidylcholine; MAG, monoacylglycerol; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PREDIMED, PREvención con Dieta MEDiterránea; PS, phosphosphingolipid; TAG, triglyceride.

of ALA and LA, particularly some seed oils, are common; hence, these fatty acids cannot be considered as highly specific markers of walnut consumption, except in populations in which the intake of seed oils is lower or has been controlled in clinical settings. In a Spanish population with dyslipidemia, higher phospholipid proportions of ALA were correlated with walnuts; the coefficients between fatty acid proportions and the corresponding calculated intakes were  $r = 0.44$ ;  $P < 0.001$  (28). Although ALA has been identified as a reliable biomarker of walnuts, the identification of other metabolites associated with walnut consumption could provide a deeper understanding of the potential biological pathways that are influenced by the consumption of walnuts, and thus metabolites reflect the inherent variation in metabolism of diet and, therefore, more closely represent biological availability. In addition, identifying a metabolite model with several metabolites could be more specific of walnut consumption rather than the use of a single metabolite. Although ALA and LA were not measured in the platforms used for the current study, the metabolite

model selected 36:4 PC as the second metabolite with the highest coefficient. 36:4 PC is a phosphatidylcholine that can be derived from both endogenous and food sources and belongs to the LA/ALA metabolic pathway (29). However, the metabolomics platforms only identified lipids by head group and total acyl carbon number and total acyl double bond content; thus, 36:4 PC cannot be confirmed as 16:0–20:4 PC, and it could also be 18:2–18:2 PC, 18:1–18:3 PC, or others.

In previous studies, urolithins—a product of ellagitannins metabolized by gut microbiota (9)—have been commonly identified in urine after walnut consumption. Although they may be used to differentiate between the intake of walnuts and nuts because they are not abundant in the latter, they have been reported after the intake of other foods high in polyphenols, such as strawberries, blackberries, and raspberries (8). Specifically, in a previous study within the PREDIMED trial, 18 urinary metabolites, including markers of fatty acid metabolism, ellagitannin-derived microbial compounds,

**TABLE 2** Pearson correlation coefficients between metabolomics signatures and consumption

Assessment	Discovery population (PREDIMED baseline)			Internal validation population (Year 1 data in PREDIMED)
	Pearson correlation with metabolomic signature (95% CI)	Total metabolites <sup>1</sup>	No. of metabolites with positive coefficients	Pearson correlation (95% CI)
			No. of metabolites with negative coefficients	
Walnuts, g/d	0.16 (0.11, 0.20) 0.13 (0.08, 0.19) <sup>2</sup>	19 9	9 9	10 0

<sup>1</sup>Metabolite coefficients obtained 10 times in the cross-validation procedure for the elastic net continuous approach. Results using the lambda.min option. PREDIMED, PREvención con Dieta MEDiterránea.

<sup>2</sup>Pearson correlation and 95% CI between self-reported walnut intake and the multimetabolite model using only metabolites with positive values.

**TABLE 3** Pearson correlation coefficients between the metabolites consistently associated with the consumption of walnuts in the PREDIMED study<sup>1</sup>

Metabolite	
40:6 PS	-0.11 (-0.16, -0.06)
16:1 LPC	-0.08 (-0.13, -0.03)
14:0 LPC	-0.08 (-0.13, -0.03)
Isocitrate	-0.08 (-0.13, -0.03)
Succinate	-0.07 (-0.12, -0.03)
Cyclohexylamine	-0.06 (-0.11, -0.01)
36:5 PE plasmalogen	-0.05 (-0.09, -0.00)
Piperine	-0.05 (-0.09, -0.00)
Sorbitol	-0.05 (-0.09, -0.00)
<i>N</i> -acetylaspatic acid	0.05 (0.00, 0.09)
22:1 MAG	0.05 (0.00, 0.09)
4-Hydroxy-3-methylacetophenone	0.05 (0.02, 0.09)
Guanine	0.07 (0.02, 0.11)
Serine	0.07 (0.02, 0.11)
Bilirubin	0.07 (0.03, 0.12)
54:6 TAG	0.08 (0.04, 0.13)
36:4 PC	0.08 (0.04, 0.13)
Biliverdin	0.09 (0.04, 0.13)
10:2 carnitine	0.11 (0.07, 0.16)

<sup>1</sup>Values are Pearson correlations and 95% CIs for the metabolites consistently associated with the consumption of walnuts. LPC, lysophosphatidylcholine; MAG, monoacylglycerol; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PREDIMED, PREvención con Dieta MEDiterránea; PS, phosphatidylserine; TAG, triglyceride.

and metabolites of the tryptophan/serotonin pathway were associated with walnut consumption (30). In the current study, and consistent with previous results using urine samples in the PREDIMED, tryptophan was also selected in the multi-metabolite model. Tryptophan is found in considerably high amounts in walnuts, and bacterial degradation of tryptophan generates indole acid derivatives, which have been identified in urine after walnut consumption (31). Whereas urine may be more reflective of short-term or acute consumption, most of the plasma metabolites that we identified as markers of walnut intake may rather reflect the intake over days to weeks, which is more meaningful when evaluating the association with chronic disease risk.

Given that walnuts are high in unsaturated fats, it is not surprising that a large proportion of the metabolites that we identified are related to PUFA and lipid metabolic pathways. The elastic net regression identified several lipids, especially glycerophospholipids and glycerols, that were positively associated with walnut consumption, including 36:4 PC, 22:1 MAG, and 54:6 triglycerides, which could reflect the fatty acid composition of walnuts. On the other hand, 2 LPCs (14:0 and 16:1) and 40:6 PS were inversely associated with walnut consumption. In the current study, carbohydrate metabolites and metabolites from the tricarboxylic acid cycle, including isocitrate, succinate, and sorbitol, were also inversely associated with walnut intake. Because the nutritional composition of walnuts consists mostly of fatty acids and protein, and the amount of carbohydrates is relatively low, it is not unexpected that carbohydrate-related metabolites were inversely associated with walnut intake.

10:2 carnitine was also selected in the model. Carnitines can be derived from diet, mainly from animal food, but can also be synthesized in the body from several precursors, including lysine and methionine (32). Previous studies have suggested

that short-chain carnitines are associated with higher risk of cardiometabolic diseases (13, 15, 33), but the role of medium-chain carnitines has not been clearly established and it is possible that they play different functions in the  $\beta$ -oxidation of fatty acids in the mitochondria (32). Of note, we acknowledge that it is possible that some of the metabolites may have been selected by chance due to the use of agnostic machine-learning approaches.

Serine was also significantly and positively associated with walnut consumption. Serine is a nonessential amino acid that can be supplied from food or synthesized by the body from several metabolites. It can be found in higher amounts in nuts, especially walnuts (34), peanuts, and almonds, as well as other foods, including soybeans and legumes. Thus, it is plausible that combined with other lipids it can be a potential marker of the consumption of walnuts and other nuts. Finally, 2 metabolites involved in porphyrin and chlorophyll metabolism have been identified in the multimetabolite model: bilirubin and biliverdin. Bilirubin is a bile pigment, and biliverdin also belongs to the class of organic compounds known as bilirubin. It has been shown that walnut consumption alters the GI microbiota and microbially derived secondary bile acids (35). However, its role in humans is not fully understood.

Previous observational studies have shown that higher consumption of walnuts is associated with a lower risk of developing T2D and CVD (5, 6). In a report based on data from the Nurses' Health Study and the Health Professionals Follow-Up Study, walnut consumption was associated with a lower risk of incident CVD after comparing highest with lowest categories (RR: 0.81; 95% CI: 0.71, 0.91) (6). Similarly, higher walnut consumption was associated with a 33% (RR: 0.67; 95% CI: 0.54, 0.82) lower risk of T2D when comparing participants who consumed  $\geq 2$  servings/wk of walnuts with those who never or rarely consumed walnuts ( $P$  for linear trend < 0.001) (5). In the current study, after adjusting for potential confounders and self-reported walnut consumption, and consistent with previous findings showing inverse associations of walnut consumption with cardiometabolic diseases, we observed that the metabolite model predicting walnut consumption was also associated with 17% (HR: 0.83; 95% CI: 0.71, 0.97) lower risk of T2D and 29% (HR: 0.71; 95% CI: 0.60, 0.85) lower risk of CVD.

The current study has several strengths, including the large sample size, detailed covariate data to control for confounding, and a metabolite profile approach that allowed the analysis of >300 metabolites. We employed agnostic machine-learning models using well-characterized metabolites. Moreover, we cross-validated our results internally in the discovery population using baseline data and conducted replication analysis using data at year 1. The current study also has limitations. First, because dietary data were collected using an FFQ, measurement errors may be present compared to use of short-term biomarkers of intake. However, the validity and reproducibility of the FFQ have been reported previously. Of note, the correlation between total nut consumption assessed by FFQ and 3-d dietary records was relatively high ( $r = 0.55$ ) (36). Second, because of the observational design of our study, we are unable to establish causality of the association between the metabolomic signatures and cardiometabolic diseases. Nevertheless, we performed rigorous multivariable adjustment to minimize residual confounding. Furthermore, although we evaluated the cross-population reproducibility of the metabolite profiles, it should be validated in independent populations. Similarly, because the study was conducted in an older Mediterranean population, the

**TABLE 4** HRs (95% CIs) for incident type 2 diabetes and cardiovascular disease according to metabolites correlated with walnut consumption in the PREDIMED study<sup>1</sup>

	PREDIMED study baseline <sup>2</sup>		PREDIMED year 1 <sup>3</sup>	
	HR (95% CI)	P	HR (95% CI)	P
Type 2 diabetes				
No. of cases/total participants	245/923		161/704	
Score of metabolites predicting walnuts				
Model 1 (basic model)	0.82 (0.72, 0.94)	0.005	0.76 (0.63, 0.91)	0.004
Model 2 (+ sociodemographic model)	0.86 (0.74, 0.99)	0.04	0.80 (0.64, 0.99)	0.044
Model 3 (+ diet model)	0.86 (0.74, 1.00)	0.05	0.82 (0.64, 1.05)	0.125
Model 4 (+ consumption of walnuts)	0.83 (0.71, 0.97)	0.02	0.81 (0.63, 1.03)	0.092
Cardiovascular disease				
No. of cases/total participants	222/993		159/916	
Model 1 (basic model)	0.72 (0.63, 0.84)	<0.001	0.74 (0.62, 0.88)	<0.001
Model 2 (+ sociodemographic model)	0.67 (0.58, 0.78)	<0.001	0.76 (0.62, 0.92)	0.007
Model 3 (+ diet model)	0.66 (0.56, 0.77)	<0.001	0.75 (0.60, 0.93)	0.011
Model 4 (+ consumption of walnuts)	0.71 (0.60, 0.85)	<0.001	0.76 (0.61, 0.95)	0.016

<sup>1</sup>Model 1 (basic model): adjusted for age, sex, and propensity scores; stratified by intervention group and recruitment center. Model 2 (sociodemographic model): model 1 + BMI, smoking status (never, former, or current smoker), alcohol intake and squared alcohol intake (grams per day), education level (primary, secondary, academic), physical activity (metabolic-equivalent minutes per day), family history of CHD (yes/no), baseline dyslipidemia or lipid-lowering medication use (yes/no), baseline hypertension or antihypertensive use (yes/no), and T2D prevalence (only in CVD analysis). Model 3: model 2 + total energy and intakes of vegetables, fruits, cereals, red and processed meat, fish, olive oil, eggs, legumes, and dairy in quintiles. Model 4: model 3 + intake of walnuts from which metabolite set was derived. CHD, coronary heart disease; CVD, cardiovascular disease; PREDIMED, PREvención con Dieta MEDiterránea; T2D, type 2 diabetes.

<sup>2</sup>Analysis of CVD risk was conducted among the 993 participants of the PREDIMED CVD case-cohort data set or 923 in the T2D case-cohort data set. Cox proportional hazard models, with Barlow weights (inverse probability weights to account for the overrepresentation of cases), were used to estimate HRs and their 95% CIs for risk of CVD. Person-time of follow-up was calculated as the interval between the baseline date and date of CVD or T2D event, death, or date of the last participant contact, whichever came first. HRs refer to 1-SD increase in correlated multimetabolite score.

<sup>3</sup>Walnuts intake, metabolic signatures, and covariates were assessed at year 1, and outcome was the incident CVD events occurred after the year 1 visit through the end of follow-up. The analytic models were the same as in the baseline analysis. A total of 916 participants were included in the CVD analyses and 704 in the T2D analyses.

intake of total nuts at baseline was 11 g/d and walnut intake was 6 g/d, which are higher than the intake in other populations; thus, the results cannot easily be extrapolated to the general population. Third, the analysis was limited to 385 targeted metabolites; thus, we cannot exclude that more biologically relevant metabolites regarding walnut intake were absent from the analytical data set and could be identified using untargeted approaches. The metabolomics approach used for quantifying lipids did not identify the specific fatty acids for each molecule; consequently, we can only provide the number of carbons and double bonds of each lipid, and thus some relevant walnut biomarkers may have been missed. Future studies are warranted to identify additional biomarkers of walnut intake. Finally, considering that the plasma metabolome reflects the overall metabolic homeostasis resulting from dietary intake and other biological processes, the metabolite profile model identified not only metabolites derived from food but also variations in the metabolism that are affected by dietary intake and other factors influencing metabolites. Therefore, the metabolite profile model associated with walnut consumption could reflect the combined effects of walnuts, substitution of other food by the consumption of walnuts, and individual metabolic responses to diet (37). Although all these aspects may contribute to the health effects of walnut consumption, the current study was not designed to differentiate between biomarkers of intake and metabolites that reflect individual metabolic responses.

In conclusion, we identified a panel of 19 plasma metabolites associated with walnut consumption. We also provided evidence that the identified walnut metabolite profile is associated with lower risk of T2D and CVD in Mediterranean individuals

at high cardiovascular risk. Together, these findings provide insights into potential biological mechanisms underlying associations between walnut consumption and cardiometabolic health.

### Acknowledgments

We thank the participants for their enthusiastic collaboration, the PREDIMED personnel for excellent assistance, and the personnel of all affiliated primary care centers. The authors' responsibilities were as follows—MG-F, PH-A, J-PD-C, MR-C, CR, ET, MAM-G, FHB, and JS-S: conceived and designed the work; DC, RE, MF, MAM-G, and JS-S: coordinated the subject recruitment at the outpatient clinics and clinical data collection; CD and CBC: conducted the metabolomics data analysis; MG-F and PH-A: conducted the statistical analysis and are the guarantors of this work and, as such, take responsibility for the integrity of the data and the accuracy of the data analysis; MG-F, PH-A, FBH, and JS-S: had access to all the data in the study; MG-F, PH-A, J-PD-C, MR-C, CR, ET, JL, CW, CBC, LL, MAM-G, FBH, and JS-S: interpreted the data; MG-F and PH-A: drafted the manuscript; and all authors: made critical revisions to the manuscript for key intellectual content and read and approved the final manuscript.

### References

1. Li L, Tsao R, Yang R, Kramer JKG, Hernandez M. Fatty acid profiles, tocopherol contents, and antioxidant activities of heartnut (*Juglans ailanthifolia* Var. *cordiformis*) and Persian walnut (*Juglans regia* L.). *J Agric Food Chem* 2007;55:1164–9.
2. Maguire LS, O'Sullivan SM, Galvin K, O'Connor TP, O'Brien NM. Fatty acid profile, tocopherol, squalene and phytosterol content of



- walnuts, almonds, peanuts, hazelnuts and the macadamia nut. *Int J Food Sci Nutr* 2004;55:171–8.
3. Baum SJ, Kris-Etherton PM, Willett WC, Lichtenstein AH, Rudel LL, Maki KC, Whelan J, Ramsden CE, Block RC. Fatty acids in cardiovascular health and disease: a comprehensive update. *J Clin Lipidol* 6:216–34.
  4. Guasch-Ferré M, Li J, Hu FB, Salas-Salvado J, Tobias DK. Effects of walnut consumption on blood lipids and other cardiovascular risk factors: an updated meta-analysis and systematic review of controlled trials. *Am J Clin Nutr* 2018;174–87.
  5. Pan A, Sun Q, Manson JE, Willett WC, Hu FB. Walnut consumption is associated with lower risk of type 2 diabetes in women. *J Nutr* 2013;143(4):512–8.
  6. Guasch-Ferré M, Liu X, Malik VS, Sun Q, Willett WC, Manson JE, Rexrode KM, Li Y, Hu FB, Bhupathiraju SN. Nut consumption and risk of cardiovascular disease. *J Am Coll Cardiol* 2017;70:2519–32.
  7. Guasch-Ferré M, Bhupathiraju SN, Hu FB. Use of metabolomics in improving assessment of dietary intake. *Clin Chem* 2018;64:82–98.
  8. Garcia-Aloy M, Hulshof PJM, Estruel-Amades S, Osté MCJ, Lankinen M, Geleijnse JM, De Goede J, Ulaszewska M, Mattivi F, Bakker SJL, et al. Biomarkers of food intake for nuts and vegetable oils: an extensive literature search. *Genes Nutr* 2019;14:7.
  9. Tomás-Barberán FA, González-Sarriás A, García-Villalba R, Núñez-Sánchez MA, Selma MV, García-Conesa MT, Espín JC. Urolithins, the rescue of “old” metabolites to understand a “new” concept: metabolotypes as a nexus among phenolic metabolism, microbiota dysbiosis, and host health status. *Mol Nutr Food Res* 2017;61(1). doi: 10.1002/mnfr.201500901.
  10. Martínez-González MA, Corella D, Salas-Salvado J, Ros E, Covas MI, Fiol M, Warnberg J, Aros F, Ruiz-Gutiérrez V, Lamuela-Raventós RM, et al. Cohort profile: design and methods of the PREDIMED study. *Int J Epidemiol* 2012;41:377–85.
  11. Estruch R, Ros E, Salas-Salvado J, Covas M-I, Corella D, Arós F, Gómez-Gracia E, Ruiz-Gutiérrez V, Fiol M, Lapetra J, et al. Primary prevention of cardiovascular disease with a Mediterranean diet supplemented with extra-virgin olive oil or nuts. *N Engl J Med* 2018;378:e34.
  12. Ruiz-Canela M, Toledo E, Clish CB, Hruby A, Liang L, Salas-Salvado J, Razquin C, Corella D, Estruch R, Ros E, et al. Plasma branched-chain amino acids and incident cardiovascular disease in the PREDIMED trial. *Clin Chem* 2016;62:582–92.
  13. Guasch-Ferre M, Zheng Y, Ruiz-Canela M, Hruby A, Martínez-González MA, Clish CB, Corella D, Estruch R, Ros E, Fito M, et al. Plasma acylcarnitines and risk of cardiovascular disease: effect of Mediterranean diet interventions. *Am J Clin Nutr* 2016;103:1408–16.
  14. Ruiz-Canela M, Guasch-Ferré M, Toledo E, Clish CB, Razquin C, Liang L, Wang DD, Corella D, Estruch R, Hernáez Á, et al. Plasma branched chain/aromatic amino acids, enriched Mediterranean diet and risk of type 2 diabetes: case-cohort study within the PREDIMED trial. *Diabetologia* 2018;61(7):1560–71.
  15. Guasch-Ferré M, Ruiz-Canela M, Li J, Zheng Y, Bulló M, Wang DD, Toledo E, Clish C, Corella D, Estruch R, et al. Plasma acylcarnitines and risk of type 2 diabetes in a Mediterranean population at high cardiovascular risk. *J Clin Endocrinol Metab* 2019;104:1508–19.
  16. Mataix J. Tablas de composición de alimentos. 4th ed. Granada (Spain): University of Granada; 2003.
  17. Moreiras O, Cabrera L, editors. Tablas de composición de alimentos (Food composition tables). 9th ed. Madrid (Spain): Ediciones Pirámide; 2005.
  18. Elosua R, Marrugat J, Molina L, Pons S, Pujol E; The MARATHOM Investigators. Validation of the Minnesota Leisure Time Physical Activity Questionnaire in Spanish men. *Am J Epidemiol* 1994;139:1197–209.
  19. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, Lewis GD, Fox CS, Jacques PF, Fernandez C, et al. Metabolite profiles and the risk of developing diabetes. *Nat Med* 2011;17:448–53.
  20. O’Sullivan JF, Morningstar JE, Yang Q, Zheng B, Gao Y, Jeanfavre S, Scott J, Fernandez C, Zheng H, O’Connor S, et al. Dimethylguanidino valeric acid is a marker of liver fat and predicts diabetes. *J Clin Invest* 2017;127:4394–402.
  21. Paynter NP, Balasubramanian R, Giulianini F, Wang DD, Tinker LF, Gopal S, Deik AA, Bullock K, Pierce KA, Scott J, et al. Metabolic predictors of incident coronary heart disease in women. *Circulation* 2018;137:841–53.
  22. Stekhoven DJ, Bühlmann P. Missforest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* 2012;28:112–8.
  23. Gromski PS, Xu Y, Kotze HL, Correa E, Ellis DI, Armitage EG, Turner ML, Goodacre R. Influence of missing values substitutes on multivariate analysis of metabolomics data. *Metabolites* 2014;4:433–52.
  24. Wei R, Wang J, Su M, Jia E, Chen S, Chen T, Ni Y. Missing value imputation approach for mass spectrometry-based metabolomics data. *Sci Rep* 2018;8 :663.
  25. Hernández-Alonso P, Papandreou C, Bulló M, Ruiz-Canela M, Dennis C, Deik A, Wang DD, Guasch-Ferré M, Yu E, Toledo E, et al. Plasma metabolites associated with frequent red wine consumption: a metabolomics approach within the PREDIMED study. *Mol Nutr Food Res* 2019;63:1900140.
  26. van den Berg RA, Hoefsloot HCJ, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* 2006;7:142.
  27. McKay DL, Chen CYO, Yeum KJ, Matthan NR, Lichtenstein AH, Blumberg JB. Chronic and acute effects of walnuts on antioxidant capacity and nutritional status in humans: a randomized, cross-over pilot study. *Nutr J BioMed Central* 2010;9:21.
  28. Sala-Vila A, Cofán M, Pérez-Heras A, Núñez I, Gilbert R, Junyent M, Mateo-Gallego R, Cénarro A, Civeira F, Ros E. Fatty acids in serum phospholipids and carotid intima-media thickness in Spanish subjects with primary dyslipidemia. *Am J Clin Nutr* 2010;92:186–93.
  29. Chong J, Wishart DS, Xia J. Using MetaboAnalyst 4.0 for comprehensive and integrative metabolomics data analysis. *Curr Protocols Bioinform* 2019;68:e86.
  30. Garcia-Aloy M, Llorach R, Urpi-Sarda M, Tulipani S, Estruch R, Martínez-González MA, Corella D, Fitó M, Ros E, Salas-Salvado J, et al. Novel multimetabolite prediction of walnut consumption by a urinary biomarker model in a free-living population: the PREDIMED study. *J Proteome Res* 2014;13:3476–83.
  31. Magnus V, Simaga S, Iskrić S, Kveder S. Metabolism of tryptophan, indole-3-acetic acid, and related compounds in parasitic plants from the genus *Orobanchae*. *Plant Physiol* 1982;69:853–8.
  32. Violante S, Ijlst L, Te Brinke H, Koster J, Tavares de Almeida I, Wanders RJA, Ventura FV, Houten SM. Peroxisomes contribute to the acylcarnitine production when the carnitine shuttle is deficient. *Biochim Biophys Acta* 2013;1831:1467–74.
  33. Sun L, Liang L, Gao X, Zhang H, Yao P, Hu Y, Ma Y, Wang F, Jin Q, Li H, et al. Early prediction of developing type 2 diabetes by plasma acylcarnitines: a population-based study. *Diabetes Care* 2016;39:1563–70.
  34. Hu Q, Liu J, Li J, Liu H, Dong N, Geng Y-Y, Lu Y, Wang Y. Phenolic composition and nutritional attributes of diaphragma juglandis fructus and shell of walnut (*Juglans regia* L.). *Food Sci Biotechnol* 2020;29:187–96.
  35. Holscher HD, Guetterman HM, Swanson KS, An R, Matthan NR, Lichtenstein AH, Novotny JA, Baer DJ. Walnut consumption alters the gastrointestinal microbiota, microbially derived secondary bile acids, and health markers in healthy adults: a randomized controlled trial. *J Nutr* 2018;148:861–7.
  36. Fernández-Ballart JD, Pinol JL, Zazpe I, Corella D, Carrasco P, Toledo E, Pérez-Bauer M, Martínez-González MA, Salas-Salvado J, Martín-Moreno JM. Relative validity of a semi-quantitative food-frequency questionnaire in an elderly Mediterranean population of Spain. *Br J Nutr* 2010;103(12):1808–16.
  37. Li J, Guasch-Ferré M, Chung W, Ruiz-canela M, Toledo E, Corella D, Bhupathiraju SN, Tobias DK, Tabung FK, Hu J, et al. The Mediterranean diet, plasma metabolome, and cardiovascular disease risk. *Eur Heart J* 2020;41:ehaa209.