

RESEARCH

Open Access



# Olive oil consumption, plasma metabolites, and risk of type 2 diabetes and cardiovascular disease

Jesús F. García-Gavilán<sup>1,2,3\*</sup>, Nancy Babio<sup>1,2,3</sup>, Estefanía Toledo<sup>3,4</sup>, Zhila Semnani-Azad<sup>5</sup>, Cristina Razquin<sup>3,4</sup>, Courtney Dennis<sup>6</sup>, Amy Deik<sup>6</sup>, Dolores Corella<sup>3,7</sup>, Ramón Estruch<sup>3,8,9</sup>, Emilio Ros<sup>3,10</sup>, Montserrat Fitó<sup>3,11</sup>, Fernando Arós<sup>12</sup>, Miquel Fiol<sup>3,13</sup>, José Lapetra<sup>3,14</sup>, Rosa Lamuela-Raventós<sup>3,9,15</sup>, Clary Clish<sup>6</sup>, Miguel Ruiz-Canela<sup>3,4</sup>, Miguel Ángel Martínez-González<sup>3,4,5</sup>, Frank Hu<sup>5,16,17</sup>, Jordi Salas-Salvadó<sup>1,2,3†</sup> and Marta Guasch-Ferré<sup>5,18,19\*†</sup>

## Abstract

**Background** Olive oil consumption has been inversely associated with the risk of type 2 diabetes (T2D) and cardiovascular disease (CVD). However, the impact of olive oil consumption on plasma metabolites remains poorly understood. This study aims to identify plasma metabolites related to total and specific types of olive oil consumption, and to assess the prospective associations of the identified multi-metabolite profiles with the risk of T2D and CVD.

**Methods** The discovery population included 1837 participants at high cardiovascular risk from the PREvención con Dieta MEDiterránea (PREDIMED) trial with available metabolomics data at baseline. Olive oil consumption was determined through food-frequency questionnaires (FFQ) and adjusted for total energy. A total of 1522 participants also had available metabolomics data at year 1 and were used as the internal validation sample. Plasma metabolomics analyses were performed using LC–MS. Cross-sectional associations between 385 known candidate metabolites and olive oil consumption were assessed using elastic net regression analysis. A 10-cross-validation (CV) procedure was used, and Pearson correlation coefficients were assessed between metabolite-weighted models and FFQ-derived olive oil consumption in each pair of training–validation data sets within the discovery sample. We further estimated the prospective associations of the identified plasma multi-metabolite profile with incident T2D and CVD using multi-variable Cox regression models.

**Results** We identified a metabolomic signature for the consumption of total olive oil (with 74 metabolites), VOO (with 78 metabolites), and COO (with 17 metabolites), including several lipids, acylcarnitines, and amino acids. 10-CV Pearson correlation coefficients between total olive oil consumption derived from FFQs and the multi-metabolite profile were 0.40 (95% CI 0.37, 0.44) and 0.27 (95% CI 0.22, 0.31) for the discovery and validation sample, respectively. We identified several overlapping and distinct metabolites according to the type of olive oil consumed. The baseline metabolite profiles of total and extra virgin olive oil were inversely associated with CVD incidence (HR per 1SD: 0.79;

<sup>†</sup>Jordi Salas-Salvadó and Marta Guasch-Ferré are senior authors.

\*Correspondence:

Jesús F. García-Gavilán  
[jesusfrancisco.garcia@urv.cat](mailto:jesusfrancisco.garcia@urv.cat)  
Marta Guasch-Ferré  
[marta.guasch@sund.ku.dk](mailto:marta.guasch@sund.ku.dk)

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

95% CI 0.67, 0.92 for total olive oil and 0.70; 0.59, 0.83 for extra virgin olive oil) after adjustment for confounders. However, no significant associations were observed between these metabolite profiles and T2D incidence.

**Conclusions** This study reveals a panel of plasma metabolites linked to the consumption of total and specific types of olive oil. The metabolite profiles of total olive oil consumption and extra virgin olive oil were associated with a decreased risk of incident CVD in a high cardiovascular-risk Mediterranean population, though no associations were observed with T2D incidence.

*Trial registration:* The PREDIMED trial was registered at ISRCTN (<http://www.isrctn.com/>, ISRCTN35739639).

**Keywords** Olive oil, Metabolomics, Cardiovascular disease, Type 2 diabetes

## Introduction

Olive oil has been traditionally used as the main culinary and dressing fat in Mediterranean regions and is gaining global popularity due to its remarkable nutritional profile and health benefits. It has been proposed as one of the key components of the Mediterranean Diet (MedDiet), which makes it cardio-protective [1]. The best nutritional and organoleptic quality olive oil, extra-virgin olive oil, and virgin olive oil (VOO) varieties are obtained by mechanically pressing olives. They contain an exceptional matrix of lipids rich in monounsaturated fatty acids (mainly oleic acid) and high amounts of bioactive compounds, including polyphenols (hydroxytyrosol and oleuropein), lipid derivatives (squalene, tocopherols), and vitamin E, and have a richer taste, color, and aroma than other common varieties. Because of its processing, the refined or common variety of olive oil (COO) has a similar lipid profile but a lower content of phytochemicals [2]. Evidence from epidemiological studies and clinical trials support that olive oil consumption is associated with a lower risk of type 2 diabetes (T2D), cardiovascular diseases (CVD), and mortality [3–5], and has demonstrated protective effects on lipid metabolism, inflammation, endothelial function, and oxidative stress [6–8]. Despite the mounting evidence supporting olive oil's beneficial impact on human health, the lack of detailed differentiation between olive oil types in existing studies is a significant limitation. This oversight hinders a comprehensive understanding of the distinctive health properties offered by VOO.

Despite the wide range of health benefits attributed to olive oil consumption, the biological mechanisms underlying these salutary effects have not been well defined. Nutritional metabolomics, an evolving approach, holds great promise in enhancing our understanding of the biological effects of nutritional factors and may also help to identify potential novel biomarkers related to dietary intake or the subsequent metabolic response to this intake. Along these lines, the distinct phytochemical concentrations found in olive oil varieties may confer specific health advantages [7]. For example, VOO consumption

has been associated with better HDL functionality, indicating a potential mediating role between metabolic pathways associated with cardiovascular health and VOO consumption [9, 10]. We hypothesized that the amount of total olive oil consumed, and specifically if the consumption is mainly from VOO or common olive oil, is associated with unique metabolite profiles and that these multi-metabolite profiles are related to a lower incidence of T2D and CVD.

In the current study, leveraging dietary and metabolomics data in the PREvención con DIeta MEDiterránea (PREDIMED) study, we used an agnostic machine learning approach to identify plasma metabolite profiles associated with total olive oil, VOO, and COO. We then evaluated whether the identified multi-metabolite profiles are associated with T2D and CVD incidence risk independently of known risk factors and diet.

## Methods

### Study population

The present analysis was conducted in the context of the PREDIMED study. This study was a Spanish multicenter randomized controlled nutritional intervention trial conducted between 2003 and 2010. The main objective was to examine the effect of the MedDiet on the primary prevention of CVD in a population with several risk factors for CVD. The full protocol of the PREDIMED study has been previously published and can be found on the study website (<http://www.predimed.es/>) [11, 12]. Study participants provided written informed consent and all the study centers have approved the protocol by their Institutional Review Boards. The PREDIMED trial is registered at ISRCTN (<http://www.isrctn.com/> ISRCTN35739639).

The participants involved in this analysis are from three nested study samples: the first study for T2D (the PREDIMED-T2D study), the second study for CVD (the PREDIMED-CVD study), and a third subset of PREDIMED participants who completed an oral glucose tolerance test (OGTT) at baseline. The first case-cohort

study consisted of 251 participants with incident T2D cases and 694 participants without T2D at baseline (overlapping participants  $n=53$  between cases and cohort participants) [13, 14]. The second case-cohort study consisted of 229 participants with incident CVD cases (a composite of myocardial infarction, stroke, and CVD mortality) and 788 participants without CVD at baseline (overlapping  $n=37$  between cases and cohort participants). The third study consisted of 132 participants. More information about these studies is available elsewhere [15, 16].

Participants who had complete metabolomics data and nutritional data at baseline (from semi-quantitative food frequency questionnaires (FFQs)) and were not duplicated (overlapping participants between databases from different metabolomics sub-studies,  $n=122$ ) were selected ( $n=1,882$ ). Additionally, participants with missing values in FFQs at baseline ( $n=11$ ), a daily energy intake lower than 500 kcal for women or 800 kcal for men and higher than 3500 kcal for women and 4000 kcal for men ( $n=30$ ) [17], or with missing values in  $\geq 20\%$  metabolites ( $n=4$ ) were excluded. Therefore, this analysis included 1,837 participants at baseline (634 participants allocated to the MedDiet supplemented with VOO group, 630 to the MedDiet supplemented with nuts, and 573 participants to the control group) (Additional file 1: Figure S1).

Further, an internal validation in the same population was conducted using dietary and metabolomics data from the 1-year visit in 1522 study participants (Additional file 1: Figure S1).

#### Dietary assessment

Dietary data were obtained using a validated 137-item semi-quantitative FFQ that trained dietitians collected from the participants at baseline and 1-year visits in face-to-face interviews [18]. Food, nutrients, and energy intake were estimated using Spanish food composition tables [19, 20]. Total olive oil, VOO, and COO consumption were derived from these FFQs. Total olive oil consumption was considered the sum of VOO and COO.

#### Anthropometric measurements and other covariates

At baseline and 1-year visits, blood pressure (in triplicate) and anthropometric measurements such as weight, height, and waist circumference were measured according to the study protocol by trained staff. Additionally, physical activity was assessed with the validated Spanish version of the Minnesota Leisure-Time Physical Activity questionnaire [21] and other information about lifestyle, medical conditions, or medication use was also collected.

#### Metabolite profiling

At baseline and 1-year visits, overnight fasting plasma EDTA samples ( $>8$  h) were collected, processed, and stored in  $-80^{\circ}\text{C}$  freezers at each recruiting center. Before metabolomics assays, case-cohort participant samples were randomized in pairs (baseline plus 1-year visit) and sent to the Broad Institute of Harvard University and the Massachusetts Institute of Technology for analysis. Metabolic profiling of the plasma samples was performed using high-throughput liquid chromatography-tandem mass spectrometry (LC-MS) techniques [22]. After quality filtration and standardization, 400 known metabolites were quantified, of which 19 metabolites were removed from the analyses (3 metabolites that were considered as internal standards (1,2-didodecanoyl-sn-glycerol-3-phosphocholine, valine-d8, and phenylalanine-d8), 7 metabolites that were drugs (acetaminophen, metronidazole, metformin, valsartan, warfarin, verapamil, atenolol) and 9 metabolites due to  $>20\%$  of missing values). The analyses were conducted with 381 known metabolites.

To quantitatively profile polar metabolites and plasma lipids, LC-MS was used as previously described [23–25]. Amino acids (AAs) and other polar metabolites were profiled with a Nexera X2 U-HPLC (Shimadzu Corp., Marlborough, MA) coupled to a Q-Exactive mass spectrometer (ThermoFisher Scientific, Waltham, MA). Metabolites were extracted from 10  $\mu\text{L}$  plasma and 90  $\mu\text{L}$  of acetonitrile/methanol/formic acid (74.9:24.9:0.2 vol:vol:vol) that contained stable isotope-labeled internal standards [valine-d8 (Sigma-Aldrich) and phenylalanine-d8 (Cambridge Isotope Laboratories)]. After centrifuging at  $9,000\times g$  for 10 min at  $4^{\circ}\text{C}$ , the samples supernatants were injected directly onto a  $150\times 2\text{-mm}$ ,  $3\text{-}\mu\text{m}$  Atlantis HILIC column (Waters). The column was eluted isocratically at a flow rate of 250 L/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. Full-scan spectra were acquired over 70–800  $m/z$ . Fatty acids and other lipids were also profiled using a Nexera X2 U-HPLC (Shimadzu Corp., Marlborough, MA) coupled to an Exactive Plus orbitrap MS (Thermo Fisher Scientific) and were extracted from 10  $\mu\text{L}$  plasma using 190  $\mu\text{L}$  of isopropanol containing 1,2-didodecanoyl-sn-glycerol-3-phosphocholine (Avanti Polar Lipids) as an internal standard. The lipid extraction (2  $\mu\text{L}$ ) was injected into a  $100\times 2.1\text{-mm}$ ,  $1.7\text{-}\mu\text{m}$  ACQUITY BEH C8 column (Waters). The column was eluted isocratically with 80% mobile-phase A of (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. For the AAs, 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. Lipids were identified using the head group, total acyl carbon numbers, and total acyl double bond content. Pairs of pooled plasma reference samples were analyzed in intervals of 20 participant samples to assess data quality and to facilitate data standardization across the analytical queue and sample batches. One sample of each pair of the pooled references functioned as a passive quality control to assess the analytical measurement reproducibility of each metabolite. The other pooled sample was used to standardize using a “nearest neighbor” approach, i.e., 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

Baseline characteristics were presented as means and standard deviations (SD) when variables were quantitative and as percentages (*n*) when variables were categorical. The metabolites associated with total olive oil, VOO, and COO consumption were selected using plasma baseline metabolomics data (i.e., discovery sample). Plasma metabolomics data at 1 year were used as validation samples. Total olive oil, VOO, and COO consumption at baseline and 1 year were adjusted for total energy intake using the residual method [17].

The statistical quality controls used with the metabolomics data were as follows: When metabolites presented missing values of less than 20% (i.e., not detectable/quantifiable concentrations or not present metabolite), they were imputed using a random forest approach (“missForest” function from the “missForest” R package) as previous publications have recommended [26–28]. Metabolites were normalized and scaled using Blom’s rank-based inverse normal transformation [29].

Linear regression models were used to assess associations between plasma baseline metabolites and total olive oil, VOO, and COO consumption. Models were adjusted for recruiting center, age, sex, smoking status (former smoker, never smoker, smoker), BMI, education, and physical activity (METs-min/day). Multiple testing correction was performed using the

Benjamini–Hochberg procedure (FDR) and reported findings with FDR *p*-Adjusted value < 0.05.

To determine the metabolomic profile associated with each exposure (total olive oil, VOO, and COO) using an agnostic metabolomics approach including knowing metabolites, Gaussian linear regression models were used with the elastic net penalty (ENR) (“caret” v 6.0–84 and “glmnet” R Package). A tenfold cross-validation (CV) approach was performed with the discovery population (PREDIMED baseline data). First, the sample was split into 10 training-validation sets (90–10%, respectively) and we performed a tenfold CV to find the optimal value of the tuning parameter ( $\lambda$ ) that results in a mean squared error within 1 SD of the minimum (minMSE). Additionally, the  $\alpha$  parameter was evaluated from 0 (i.e., a Ridge regression) to 1 (i.e., a Lasso regression) in 0.1 increments to test the best parameters for these analyses. The best-predicting accuracy in the validation sets was obtained with  $\alpha=0.4$  for total olive oil, 0.2 for VOO, and 0.3 for COO models. After evaluating the tuning parameters, a tenfold CV was performed again with the discovery population, the coefficients from each tenfold CV iteration were extracted and were constructed weighted models using the regression coefficients of the selected metabolites from each ENR in the training set to validate with its validation data set pair. Pearson correlation coefficients were determined between total olive oil, VOO, and COO consumption and the metabolomics profile in the pair of training validation data sets in both the discovery and internal validation samples (i.e., PREDIMED baseline data and PREDIMED 1-year data). For reproducibility, regression coefficients were reported using 10 iterations of the tenfold CV elastic net regression in the entire data set. These analyses were based on consistency among CV runs; therefore, any *P*-value was derived.

The associations between the identified metabolite profiles of total olive oil, VOO, and COO consumption (1 SD) and T2D risk (245 events at baseline and 161 incident events at 1 year) within the T2D nested case-cohort study and CVD risk (222 events at baseline and 159 incident events at 1 year) within the CVD nested case-cohort study were run with Cox regressions with Barlow weights and robust variance estimator. Four multivariate models were assessed. For the baseline analysis, the first model was adjusted for age (years), sex, and propensity scores, as previously described [12], and was stratified by intervention group and recruitment center. The second model was additionally adjusted for BMI ( $\text{kg}/\text{m}^2$ ), smoking status (current, former, or never), alcohol consumption in g/day (and adding a quadratic term), educational level (primary, secondary, or college), physical activity (METs/min/day), family history of

CVD (yes/no), the baseline prevalence of dyslipidemia and hypertension, and dyslipidemia and hypertension medication. The third model included all covariables from model 2 in addition to the consumption of vegetables, fruits, cereals, nuts, eggs, legumes, fish, meat, and dairy products (all g/day). The fourth model included all covariates from model 3 in addition to the baseline consumption of total olive oil, VOO, or COO from which the metabolite set was derived, respectively. For the 1-year analyses, we used the same models as at baseline excluding those T2D or CVD cases diagnosed during the first year. Interactions between olive oil profiles and intervention groups were evaluated using the likelihood ratio test including the interaction product terms as covariables.

Several sensitivity analyses were conducted. First, we performed a principal component analysis (PCA) using the mean ENR coefficients from the metabolites consistently selected (i.e., 10 times) in each olive oil profile. A zero value was assigned whenever a particular metabolite was not found by a specific approach. Coefficients were

centered and scaled before PCA analysis. Second, we evaluated the specificity of each metabolite profile using Pearson correlations between the consumption of each type of olive oil with each metabolomics profile. Third, we conducted stratified analyses by intervention group. Fourth, we further adjusted the multi-variable Cox regression models for coffee and tea intake.

All analyses were considered statistically significant when  $P < 0.05$  and were performed using R version 4.2.2 statistical software (R Foundation for Statistical Computing).

## Results

### Characteristics of the study participants

The baseline characteristics of the study participants according to tertiles of energy-adjusted total olive oil consumption are shown in Table 1. The mean  $\pm$  SD consumption of total olive oil at baseline was  $22 \pm 8$  g/day in the lowest tertile compared to  $56 \pm 10$  g/day in the highest tertile. For VOO and COO, respectively, the

**Table 1** Baseline characteristics of study participants according to tertiles of energy-adjusted total olive oil consumption

	Tertile 1 (n = 613)	Tertile 2 (n = 612)	Tertile 3 (n = 612)
Age (years)	67 $\pm$ 6	67 $\pm$ 6	67 $\pm$ 6
Women, n (%)	339 (55)	363 (59)	356 (58)
Body mass index (kg/m <sup>2</sup> )	30 $\pm$ 3	30 $\pm$ 4	30 $\pm$ 4
Waist circumference (cm)	101 $\pm$ 10	100 $\pm$ 10	101 $\pm$ 10
Type 2 diabetes, n (%)	178 (29)	165 (27)	192 (31)
Hypercholesterolemia, n (%)	485 (79)	464 (76)	462 (76)
Hypertension, n (%)	531 (87)	540 (88)	532 (87)
Family history of CVD, n (%)	152 (25)	145 (24)	154 (25)
Current smoking, n (%)	103 (17)	101 (17)	83 (14)
Vegetables (g/day)	343 $\pm$ 164	323 $\pm$ 149	328 $\pm$ 136
Fruit (g/day)	374 $\pm$ 215	356 $\pm$ 183	352 $\pm$ 193
Legumes (g/day)	22 $\pm$ 16	20 $\pm$ 11	19 $\pm$ 10
Cereals (g/day)	249 $\pm$ 104	236 $\pm$ 106	208 $\pm$ 88
Dairy (g/day)	402 $\pm$ 235	376 $\pm$ 222	348 $\pm$ 202
Total meat (g/day)	139 $\pm$ 60	136 $\pm$ 55	127 $\pm$ 52
Total fish (g/day)	99 $\pm$ 50	104 $\pm$ 64	101 $\pm$ 43
Total olive oil (g/day)	22 $\pm$ 8	40 $\pm$ 12	56 $\pm$ 10
Virgin olive oil (g/day)	9 $\pm$ 11	21 $\pm$ 20	36 $\pm$ 27
Common olive oil (g/day)	12 $\pm$ 12	18 $\pm$ 19	19 $\pm$ 25
Nuts (g/day)	12 $\pm$ 14	11 $\pm$ 14	10 $\pm$ 13
Wine (g/day)	70 $\pm$ 115	75 $\pm$ 132	55 $\pm$ 99
Alcohol (g/day)	10 $\pm$ 16	11 $\pm$ 18	8 $\pm$ 12
Total energy (kcal/day)	2297 $\pm$ 489	2309 $\pm$ 670	2243 $\pm$ 450
Adherence to the MedDiet	8 $\pm$ 2	9 $\pm$ 2	9 $\pm$ 2

Values are means  $\pm$  standard deviation for continuous variables or number (%) for categorical variables

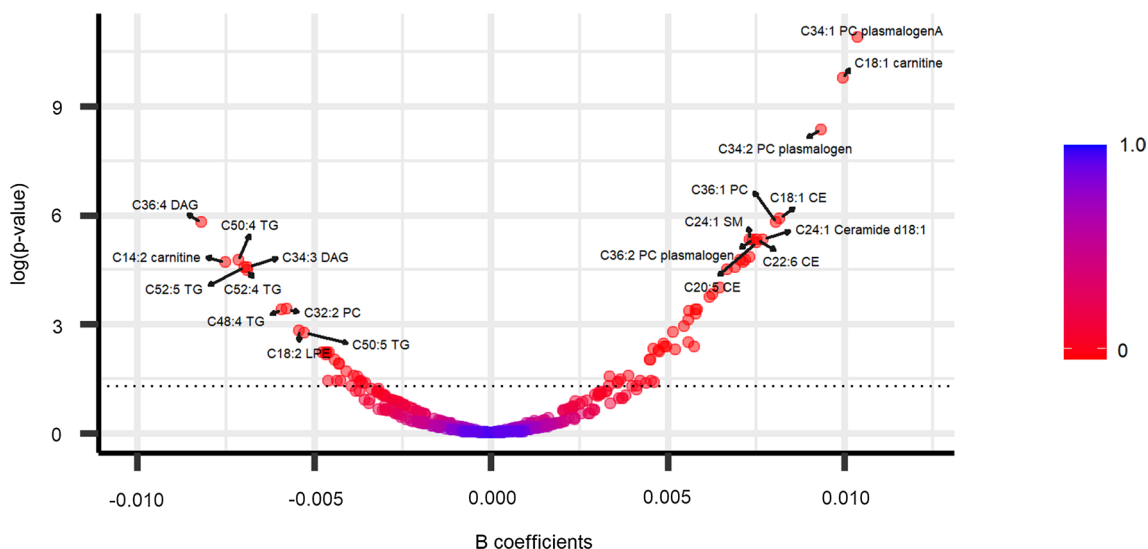
CVD Cardiovascular disease, MedDiet Mediterranean diet

Total olive oil consumption was adjusted for total energy intake using the residual method

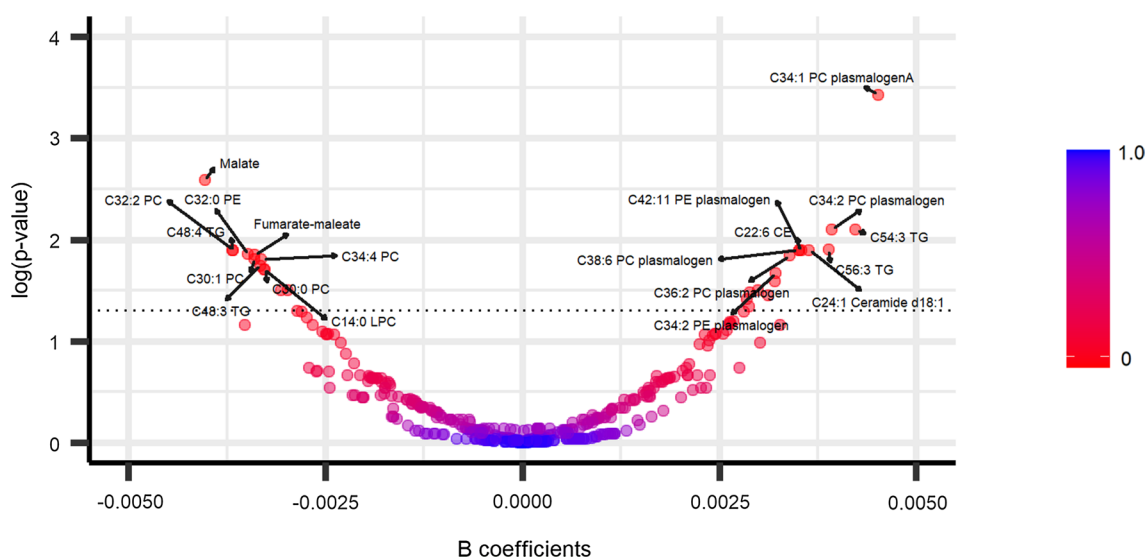
mean consumption was  $9 \pm 11$  g/day and  $12 \pm 12$  g/day in the first tertile, and  $36 \pm 27$  g/day and  $19 \pm 25$  g/day in the third tertile. In the top tertile, a higher percentage of participants had T2D, were less likely to smoke, and consumed lower amounts of fruits, vegetables, legumes, cereals, meat, and wine, as compared to participants in the lowest tertile.

### Identification of metabolites associated with olive oil consumption

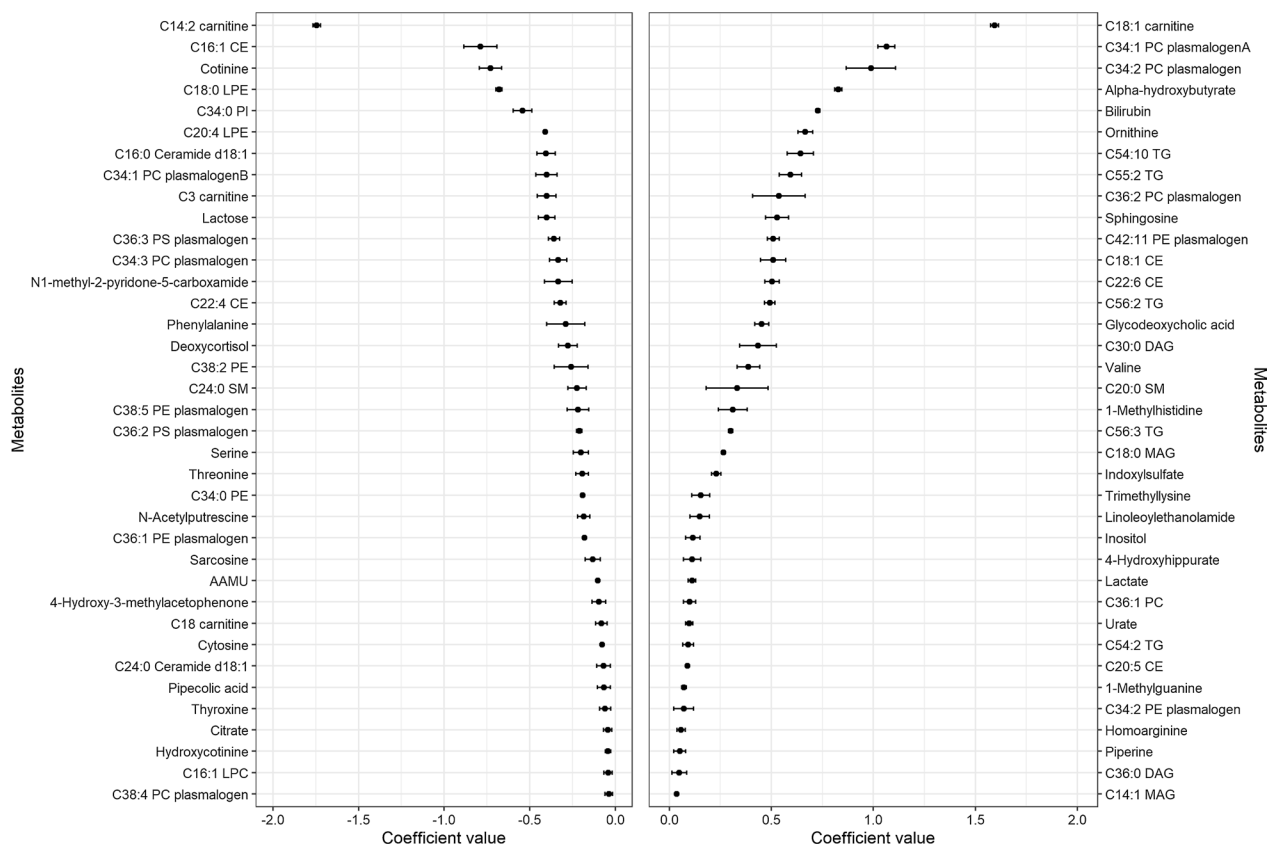
Cross-sectional associations between baseline plasma metabolites and baseline total olive oil and VOO consumption are shown in Figs. 1 and 2, and Additional file 1: Table S1. Associations between baseline plasma metabolites and baseline common olive oil consumption are shown in Additional file 1: Figure S2 and Additional file 1: Table S1. 78 metabolites were individually



**Fig. 1** Volcano plot showing the associations between plasma metabolites and total olive oil consumption at baseline. The models were adjusted by recruiting center, smoking status (former smoker, never smoker, smoker), sex, BMI, age, education, and physical activity (METs/day). An FDR < 0.05 was considered statistically significant (up dotted line)



**Fig. 2** Volcano plot showing the associations between plasma metabolites and virgin olive oil consumption at baseline. The models were adjusted by recruiting center, smoking status (former smoker, never smoker, smoker), sex, BMI, age, education, and physical activity (METs/day). An FDR < 0.05 was considered statistically significant (up dotted line)



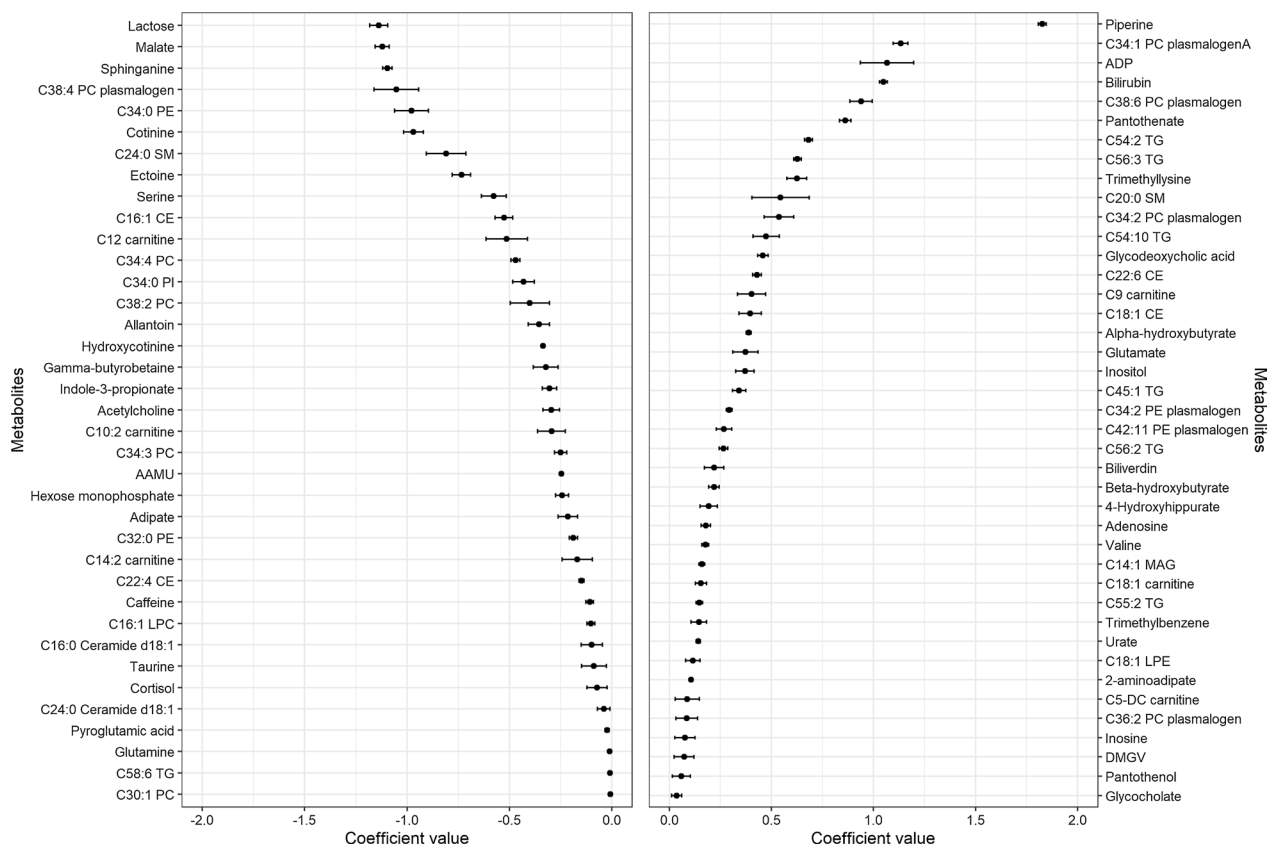
**Fig. 3** Coefficients (mean and SD) for the metabolites were selected ten times in the 10-cross validation of the continuous elastic regression for energy-adjusted total olive oil consumption. The sets of metabolites were selected using elastic continuous regression models (with lambda.min) employing the whole dataset of subjects ( $n = 1833$ ). Negative coefficients are plotted on the left, whereas positive coefficients are shown on the right

significantly associated with total olive oil consumption, 28 metabolites were significantly associated with VOO consumption, and 5 metabolites were significantly associated with COO consumption ( $FDR < 0.05$ ).

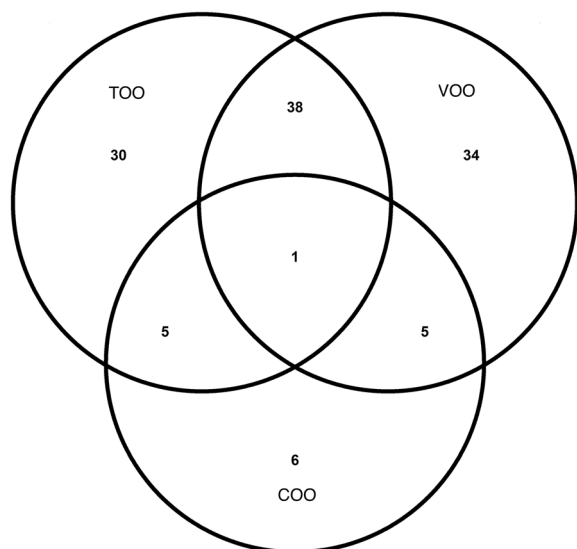
Figures 3 and 4 and Additional file 1: Figure S3 show the coefficients (mean and SD) for the ten times selected metabolites in the 10-cross validation of the continuous elastic regression for total olive oil, VOO, and COO. For total olive oil, the metabolites with the strongest inverse associations were C14:2 carnitine, C16:1 cholesterol ester (CE), and cotinine. For VOO, lactose, malate, and sphinganine exhibited the highest negative correlation coefficients; for COO, the strongest inverse associations were found for piperine, pantothenate, and adenosine diphosphate (ADP). For total olive oil, the strongest positive associations were observed for C18:1 carnitine, C34:1 PC plasmalogen A, and C34:2 PC plasmalogen; for VOO, piperine, C34:1 PC plasmalogen A, and ADP showed the strongest direct associations; and for COO, malate, sphingosine, and ornithine were the metabolites with the strongest direct associations. Only piperine was selected

in the three metabolomic profiles. This metabolite was positively associated with total olive oil and VOO, but not with COO. Figure 5 shows Venn diagrams showing the number of overlapping and different metabolites for total and subtypes of olive oil identified using elastic net continuous regressions. Thirty metabolites were selected only in the total olive oil profile, thirty-four were only in the VOO profile, and six were only selected in the COO profile.

Furthermore, to identify the principal components consisting of metabolites most associated with total olive oil, VOO, and/or COO, we also performed a PCA using the coefficients of the metabolites selected by the different olive oil consumption profiles with ENR (Additional file 1: Figure S4). The first principal component (1PC) accounted for 53% of the variability, while the second principal component (2PC) accounted for 36.3% of the variability. 1PC differentiated total olive oil and VOO profiles of the COO profile while 2PC showed differences between total olive oil and COO profiles versus VOO profile. In the PCA biplot (Additional file 1: Figure S4),



**Fig. 4** Coefficients (mean and SD) for the metabolites were selected ten times in the 10-cross validation of the continuous elastic regression for energy-adjusted virgin oil. The sets of metabolites were selected using elastic continuous regression models (with lambda.min) employing the whole dataset of subjects (n = 1833). Negative coefficients are plotted on the left, whereas positive coefficients are shown on the right



**Fig. 5** Venn diagram showing the overlapping and different metabolites for total and subtypes of olive oil consumption were identified using elastic net continuous regressions. *TOO* total olive oil, *VOO* virgin olive oil, *COO* common olive oil

we observed clusters of metabolites clustered close to the three different profiles.

Table 2 shows the Pearson correlation coefficients between each consistently selected metabolite and the consumption of total olive oil, VOO, and COO in the PREDIMED baseline data (discovery population) and 1-year data (validation sample). The Pearson correlations between the metabolite profiles and energy-adjusted olive oil consumption derived from the FFQ at baseline were 0.40 (95% CI: 0.37, 0.44) for total olive oil consumption, 0.37 (95% CI: 0.33, 0.41) for VOO consumption, and 0.25 (95% CI: 0.21, 0.29) for COO consumption. At 1 year, the Pearson correlations were 0.27 (95% CI: 0.22, 0.31) for total olive oil, 0.23 (95% CI: 0.18, 0.28) for VOO, and 0.16 (95% CI: 0.12, 0.21) for COO. Total olive oil consumption was associated with 74 metabolites, VOO with 78, and COO with 17. Additional file 1: Figure S5 shows a correlation plot between the FFQ-derived consumption of total and subtypes of olive oil and metabolite profiles.



**Table 2** Pearson correlation coefficients between metabolomics signatures and olive oil consumption

Assessment	Baseline visit			1-year visit
	Pearson correlation (95% CI) <sup>1</sup>	Total metabolites <sup>2</sup>	Metabolites with positive coefficients	Metabolites with negative coefficients
Total olive oil	0.40 (0.37, 0.44)	74	37	37
Extra virgin olive oil	0.37 (0.33, 0.41)	78	41	37
Common olive oil	0.25 (0.21, 0.29)	17	8	9

<sup>1</sup> The Pearson's coefficients reflect the correlation between FFQ-derived olive oil consumption and predicted olive oil consumption based on the olive oil-specific multi-metabolite model identified within the discovery cohort

<sup>2</sup> Number of metabolites obtained 10 times in the tenfold cross-validation procedure for the elastic net continuous regression, using the lambda.min option

Olive oil consumption was adjusted for total energy intake by the residual method

### Associations between the identified metabolomic profiles of olive oil consumption and the risk of T2D and CVD

Table 3 shows the prospective associations between the identified metabolomic profiles of total and subtypes of olive oil consumption and the incidence of T2D and CVD. No significant associations were observed for baseline and 1-year olive oil metabolite profiles and T2D risk. The identified baseline metabolomic profiles of total and subtypes of olive oil showed significant associations with CVD incidence. After adjusting for lifestyle and dietary risk factors, the HR for CVD and 95%CI (for every 1 SD increase) was 0.79 (0.67, 0.92; P-value=0.003) for the total olive oil metabolite profile, and 0.70 (0.59, 0.83; P-value=<0.001) for the VOO, but 1.37 (1.15, 1.63; P-value=<0.001) for the COO. In sensitivity analysis, additionally adjusting for coffee and tea consumption, the results remained consistent. Only the 1-year VOO metabolite profile was inversely associated with CVD risk (0.81; 0.65, 1.00; P-value=0.049).

Interactions between olive oil metabolomic profiles and intervention groups were significant in the CVD models ( $P$ -values<0.05) and non-significant in the T2D models ( $P$ -values>0.05). In the stratified analysis by intervention group (Additional file 1: Tables S2), no significant associations were observed between the olive oil metabolomic profiles and T2D incidence for any of the study intervention groups. However, inverse associations between the metabolite profile of total olive oil and VOO at baseline and the risk of developing CVD were stronger in participants allocated to the MedDiet+VOO group (Additional file 1: Table S3). A direct association between the metabolite signature of COO and a higher risk of CVD was observed. In those participants allocated to the MedDiet+nuts group, only the COO profile showed a positive significant association with CVD incidence. In participants allocated to the control group, an inverse association was observed for the VOO metabolite profile and risk of CVD. At 1 year, only in those participants allocated to the MedDiet+VOO group,

a significant inverse association between the total olive oil, and VOO metabolomics signature and CVD was observed, and the COO metabolomic profile was directly associated with CVD risk (Additional file 1: Table S3).

### Discussion

Leveraging the integrated dietary and metabolomics data in a Mediterranean population at high cardiovascular disease risk, using an agnostic machine-learning approach, we identified a metabolomic signature for the consumption of total olive oil (with 74 metabolites), VOO (with 78 metabolites), and COO (with 17 metabolites). The metabolomics signature included several lipids, acylcarnitines, and amino acids. Notably, we identified several overlapping and distinct metabolites according to the type of olive oil consumed. Among the significant findings, certain lipids, including plasmalogens, triacylglycerol (TAG), and the organic acid ADP, exhibited positive associations with total olive oil and virgin olive oil (VOO), while no such correlations were observed with common olive oil (COO). Furthermore, we observed a strong and consistent positive correlation between the multi-metabolite profiles and olive oil consumption as assessed through food-frequency questionnaires (FFQ) at both baseline and one year. Linking these metabolites to disease risk, the metabolite profiles associated with total olive oil and VOO showed inverse associations with CVD risk after adjusting for sociodemographic and dietary factors including FFQ-derived olive oil consumption. However, no significant associations between the multi-metabolite profiles and T2D were observed. To the best of our knowledge, this is the first study evaluating the association between plasma metabolite profiles of total and specific types of olive oil and the risk of cardiometabolic diseases. Overall, our findings provide novel insights into the health-promoting benefits of olive oil consumption and highlight the potential relevance of olive oil-related metabolites in relation to chronic disease

**Table 3** Hazard ratios (95% CIs) for incident type 2 diabetes and cardiovascular diseases according to multi-metabolite associated with 1 SD of olive oil consumption in the PREDIMED study

Type 2 diabetes				
	Baseline visit <sup>1</sup>		1-year visit <sup>2</sup>	
	HR (95% CI)	P	HR (95% CI)	P
Cases/total participants	245/923		161/704	
Total olive oil				
Model 1	1.01 (0.85, 1.19)	0.949	1.00 (0.79, 1.27)	0.998
Model 2	1.10 (0.91, 1.32)	0.340	1.04 (0.80, 1.36)	0.765
Model 3	1.11 (0.91, 1.35)	0.296	1.12 (0.84, 1.50)	0.448
Extra virgin olive oil				
Model 1	0.99 (0.82, 1.18)	0.882	0.87 (0.69, 1.09)	0.228
Model 2	1.08 (0.88, 1.33)	0.456	0.90 (0.70, 1.16)	0.415
Model 3	1.08 (0.87, 1.35)	0.479	0.98 (0.74, 1.28)	0.863
Common olive oil				
Model 1	1.01 (0.85, 1.21)	0.882	1.16 (0.89, 1.51)	0.228
Model 2	0.97 (0.79, 1.18)	0.735	1.15 (0.86, 1.54)	0.354
Model 3	0.96 (0.78, 1.19)	0.731	1.06 (0.78, 1.45)	0.700
Cardiovascular disease				
Cases/total participants	222/993		159/916	
Total olive oil				
Model 1	0.77 (0.67, 0.89)	<0.001	0.90 (0.76, 1.06)	0.194
Model 2	0.77 (0.66, 0.89)	<0.001	0.89 (0.74, 1.06)	0.178
Model 3	0.79 (0.67, 0.92)	0.003	0.85 (0.69, 1.05)	0.135
Extra virgin olive oil				
Model 1	0.69 (0.59, 0.80)	<0.001	0.83 (0.70, 0.99)	0.033
Model 2	0.69 (0.58, 0.80)	<0.001	0.81 (0.67, 0.97)	0.022
Model 3	0.70 (0.59, 0.83)	<0.001	0.81 (0.65, 1.00)	0.049
Common olive oil				
Model 1	1.41 (1.20, 1.65)	<0.001	1.18 (0.97, 1.45)	0.099
Model 2	1.39 (1.18, 1.65)	<0.001	1.17 (0.95, 1.45)	0.136
Model 3	1.37 (1.15, 1.63)	<0.001	1.17 (0.93, 1.48)	0.184

Model 1: adjusted for age (years), sex, and propensity scores; stratified by intervention group and recruitment center. Model 2: model 1 + BMI, smoking status (never, former, or current smoker), alcohol intake and squared alcohol intake (g/day), education level (primary, secondary, academic) physical activity (metabolic-equivalent minutes per day), family history of CHD (yes/no), dyslipidemia, hypertension, and dyslipidemia and hypertension treatment. Model 3: model 2 + consumption of vegetables, fruits, cereals, nuts, eggs, legumes, fish, meat, and dairy (g/day). Abbreviations: HR, hazard ratio; CI, confidence interval; CVD, cardiovascular disease; T2D, type 2 diabetes; BMI, body mass index

<sup>1</sup> Analysis of T2D risk was conducted in the 923 participants from the PREDIMED-T2D case-cohort database and the analysis of CVD risk was conducted in the 993 participants from the PREDIMED-CVD case-cohort database. Cox proportional hazard models with Barlow weights were used to estimate HRs and their 95% CIs for T2D. Person-time of follow-up was calculated as the interval between the baseline data and the date of T2D or CVD event, death, or date of the last participant contact, whichever came first. HRs refers to a 1-SD increase in correlated multi-metabolite scores

<sup>2</sup> Total olive oil, extra virgin olive oil, and common olive oil metabolic signatures, and covariates were assessed in the first year. The outcome was the incident T2D or CVD events occurred after the first-year visit through to the end of follow-up. The models were the same as in the baseline models. 704 participants for T2D and 916 participants for CVD were included in the analyses

Olive oil consumption variables were adjusted for total energy intake by the residual method

risk. These results pave the way for further research on the specific metabolic pathways impacted by olive oil consumption and its implications for preventing cardio-metabolic diseases.

Olive oil contains a wide variety of minor phytochemicals, such as tocopherols, carotenoids, and phenolic compounds, with recognized biological activity. Compared to common olive oil (a mixture of refined with a minor quantity of virgin olive oil), VOO is obtained exclusively by physical procedures, such as first-pressing and centrifugation, preserving a large part of its phenolic molecules, mainly hydroxytyrosol, oleocanthal, and oleuropein [30, 31]. Several studies identified a dose-dependent relationship between hydroxytyrosol in plasma or urine and VOO consumption [32, 33]. This metabolite was used as a biomarker of compliance with the intervention (MedDiet + VOO) in the PREDIMED study [12] and its biological metabolite (homovanillyl alcohol) was associated with a lower risk of CVD and total mortality [34]. In the EPIC cohort, urinary hydroxytyrosol was also correlated with total olive oil intake [35]. Unfortunately, our metabolomic platform does not allow us to determine plasma phenolic compounds. In addition, hydroxytyrosol can also be obtained from the endogenous hydroxylation of tyrosol, another common phenolic compound present in beer and wine [36, 37]. The determination of metabolite profiles of total olive oil, VOO, or COO consumption may reflect more specifically the variety of olive oil consumed and the metabolic pathways implicated after its consumption.

In previous analysis within the context of the PREDIMED Study, which evaluated the overall Mediterranean dietary pattern [38], several metabolites selected in the signature were found to be associated with olive oil consumption. Specifically, the consumption of olive oil, measured by the Mediterranean Diet Adherence Score (MEDAS), showed positive correlations with various lipids (C20:5 CE, C22:5 CE, some plasmalogens, and C24:1 SM), C18:1 carnitine and 4-pyridoxate, and negative correlations with other lipid molecules including TAGs, PC, CE or DAG, C4 and C14:2 carnitines, glycine, and pyroglutamic acid. In the present study, we also observed some of the previously mentioned associations in relation to different types of reported olive oil consumption. For instance, C20:5 CE, C18:1 carnitine, and various PCs were positively associated with VOO and/or total olive oil consumption, while C14:2 carnitine or C20:4 LPE were associated negatively. Interestingly, our findings indicate that there was only one metabolite on this list that showed an association with COO consumption. This could be attributed to the relatively low overall

consumption of this type of olive oil in our study population when compared to VOO.

While the majority of the metabolites identified in this analysis are involved in internal metabolism, it is noteworthy that some of them originate from food sources or result from microbial activity in the gastrointestinal tract. For instance, piperine showed a positive association with total olive oil and VOO consumption, while exhibiting a negative association with COO consumption. Piperine is an alkaloid found in high concentrations in black pepper [39], which happens to be one of the most commonly used spices in Mediterranean cuisine [40]. Likewise, another interesting finding was the negative association between VOO consumption and ectoine. Ectoine is a secondary metabolite [41, 42] produced by certain bacterial genera, such as *Streptomyces* spp., and has been identified in small concentrations in the human intestinal microbiota [43]. Its presence in the context of olive oil consumption suggests a potential interaction between the gut microbiome and dietary patterns. Other metabolites derived from intestinal microbiota such as 4-hydroxyhippuric acid, indole-3-propionate, and TMAO were positively related to VOO consumption, and have been associated with the metabolism of polyphenols and the consumption of foods and beverages rich in polyphenols in other studies [44, 45]. In previous studies, indole-3-propionate and TMAO were found to have negative associations with VOO and COO consumption, respectively [46]. Additionally, hydroxy cotinine and cotinine (both nicotine derivatives [47], as well as caffeine, AAMU, and N1-metil-2-piridona-5-carboxamide (caffeine derivatives [48, 49]) showed negative associations with total olive oil consumption and/or VOO. These findings may reflect the lower prevalence of smokers and coffee consumers among participants who consume VOO.

Olive oil contains mainly MUFA in the form of oleic acid [50]. Many of the identified metabolites are related to MUFA and lipid metabolic pathways, reflecting the composition of olive oil. Only five lipid metabolites (C14:0 CE, C40:6 PS, C18:2 LPC, C34:5 PC plasmalogen, and C54:5 TAG) were exclusively related to the consumption of COO. C5, C9, and C18:1 carnitines were positively associated with VOO consumption, while C10:2, C12, and C14:2 carnitines were negatively associated. Elevated concentrations of acylcarnitines may be a product of the dysregulation of fatty acid oxidation and mitochondrial function. In the PREDIMED study, medium- and long-chain acylcarnitines have been previously associated with an increased risk of CVD [15], but our results indicated that MedDiet interventions may mitigate the adverse

associations shown between higher concentrations of acylcarnitines and CVD.

Among the metabolites related to energy and carbohydrate metabolism, fructose 6-phosphate and inositol were found in the VOO profile, citrate in the total olive oil profile, lactate in both COO and total olive oil profiles, and malate in both VOO and COO profiles. Given that olive oil primarily consists of fatty acids, it is not surprising that only a few carbohydrate-related metabolites were associated with olive oil consumption. Similar observations have been reported in animal studies [51] which showed that diets rich in fats, like the Mediterranean Diet, can influence the Randle cycle, leading to increased malonyl-CoA production from  $\beta$ -oxidation, which serves as a substrate for the TCA cycle and/or gluconeogenesis [52]. Plasma glycocholate, glycodeoxycholic acid, bilirubin, and biliverdin levels are synthesized conjugated bile acids that have been positively associated with total olive oil and/or VOO in our study. Olive oil acts on the gallbladder providing its complete emptying, stimulating the synthesis of bile salts in the liver, and increasing the hepatic secretion of cholesterol [53], thus potentially explaining the association found in our study.

Our signatures have also identified several metabolites related to purine pathways. In our study, 1-methylguanine, urate, indoxyl sulfate (a uremic solute), inosine, adenosine, and ADP were positively associated with VOO or total olive oil consumption. It has been suggested that higher levels of urate in subjects with CVD may represent a compensatory response to counteract oxidative stress [54]. Another oxidative stress marker such as allantoin was negatively associated with VOO but positively associated with common olive oil [55], which suggests that VOO has a key role in these pathways with an antioxidant effect. Some metabolites were identified only in the VOO signature as adipate, acetylcholine,  $\gamma$ -butyrobetaine, 2-aminoadipate, DMGV, pantothenol or trimethyl benzene, and other metabolites only with total olive oil consumption as cytosine or pipercolic acid. We did not find an explanation in the existing literature for why these metabolites showed associations with VOO but not with COO.

These unexplained associations between specific metabolites and VOO or total olive oil consumption could potentially be attributed to the complex interactions between dietary components and individual metabolic responses.

Some metabolites that are part of the identified metabolite profiles have previously been associated with CVD and T2D [56, 57], potentially explaining the positive effects that have been reported for olive oil consumption, especially VOO, on cardiometabolic health

[1]. For example, C24:0 ceramide or  $\alpha$ -hydroxybutyrate have previously been associated with insulin resistance and increased risk of T2D, and C16:1 CE, several acylcarnitines, cortisol, or deoxycortisol (intermediate of cortisol) with increased CVD risk [56–60]. All these metabolites have been inversely associated with VOO consumption in the signature. However, other lipids such as C54:2 TAG, and C36:2 PC plasmalogen, which have also been positively associated with the consumption of total olive oil or VOO, have been associated with an increased risk of CVD and T2D. After adjusting for potential confounders, we found that the total olive oil and VOO metabolic profiles were associated with a 21% and 26% lower CVD risk, respectively, whereas the COO metabolic profile was associated with a 26% higher risk of CVD risk. However, after 1-year, no statistically significant associations between the total olive oil and COO metabolomic profiles and CVD were observed, similar to the previous analysis of the PREDIMED [61]. In the context of a MedDiet, VOO has been demonstrated to improve lipid profile, markers of glucose and inflammation, and decreased blood pressure, all considered CVD risk factors [1, 62–64]. VOO is known to have a higher concentration of polyphenols compared to COO, and these bioactive compounds are associated with various health benefits, including antioxidant and anti-inflammatory effects or improvements in lipid profile [2]. These observed differences in varying polyphenol content between olive oil subtypes may potentially explain the differences between the metabolomic profiles of VOO and COO and the risk of CVD. Additionally, the statistical differences between VOO and COO profiles could be influenced by the consumption patterns of the study participants, where the lower consumption of COO compared to the consumption of VOO implies an opposite distribution of CVD events. Curiously, similar significant results were seen in the MedDiet groups but not in the control group when stratified by intervention groups. On the other hand, no significant associations were observed between the metabolomic signatures and T2D, and some controversial findings have been reported in the literature for the associations between olive oil consumption and T2D. In the PREDIMED study, while participants allocated to the MedDiet + VOO had a lower risk of developing T2D, the consumption of olive oil alone has not been associated with T2D [65, 66]. Reverse causality might explain these findings. It is possible that participants with higher olive oil consumption, leading to higher T2DM prevalence, had a better overall lifestyle and diet compared to others in the study, influencing the observed outcomes. This highlights the complexity of dietary research and the need for careful consideration of confounding factors.

These findings need to be interpreted in the context of some limitations. First, the identified metabolite profiles are not an objective biomarker of olive oil per se, but they reflect the overall homeostasis associated with olive oil consumption, the substitution of other food by the consumption of olive oil, and the individual biological responses to diet. In addition, since olive oil consumption is usually accompanied by other foods, some of the selected metabolites may be associated with the consumption of other foods. Second, the metabolite profiles were derived from a pool of 385 annotated metabolites, while thousands of unique metabolites have been identified to date. We cannot exclude that more biologically relevant metabolites regarding olive oil intake were absent from our data set. For example, the metabolomics approach used for quantifying lipids did not identify the specific fatty acids for each molecule (we can only provide the number of carbons and double bonds of each lipid), and measures of polyphenols and phytochemicals are not available; consequently, some relevant olive oil biomarkers may have been missed. Therefore, the specificity of the identified metabolite profiles of olive oil intake remains uncertain. Future studies are warranted to identify additional objective biomarkers of olive oil intake, including urinary metabolites from dietary intake of phenolic compounds that will be assessed in the future.

Third, due to the use of FFQs for collecting dietary data, measurement errors may be present compared to the 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 olive oil intake assessed by our FFQ and 3-d dietary records was relatively high ( $r=0.60$ ) [18]. Because of our study's observational design, we cannot establish the causality of the association between the metabolomic signatures and cardiometabolic diseases. Nevertheless, we performed a rigorous multivariable adjustment to minimize residual confounding. Further, although we evaluated the cross-population reproducibility of the metabolite profiles, it should be validated in independent populations. Our population was a Mediterranean population with a high consumption of VOO even before being randomized in the study. Therefore, we cannot be ruled out reverse causation in participants with diabetes at baseline.

The current study also has several strengths. The PREDIMED study is an ideal setting to identify metabolite profiles associated with olive oil consumption because the consumption of olive oil at baseline (mean total olive oil at baseline in this population is 40 g/d) is much higher than in other populations. In addition, we

were able to differentiate between subtypes of olive oil, which have different nutritional compositions, this is a limitation of other existing studies where detailed data on specific subtypes of olive oil and comprehensive metabolomics data are not available. The present study has a large sample size and detailed covariate data to control for confounding and well-defined outcomes. In addition, we used agnostic machine learning models using more than >350 well-annotated metabolites. We cross-validated our results internally in the discovery population using baseline data and conducted replication analysis using data at year 1.

## Conclusions

In summary, our study analyzed 385 candidate metabolites and identified distinct panels associated with the consumption of total olive oil, VOO, and COO. Specifically, we identified a metabolomic signature associated with the consumption of total olive oil (with 74 metabolites), VOO (with 78 metabolites), and COO (with 17 metabolites). Moreover, our findings highlight the significance of the VOO metabolite profile in reducing the risk of CVD in Mediterranean individuals with high cardiovascular risk. These novel insights provide potential biomarkers of olive oil consumption (especially VOO) and offer valuable information on the mechanisms underlying the relationship between olive oil consumption and cardiometabolic diseases, including CVD and type 2 diabetes (T2D). By advancing our understanding of the metabolic responses to olive oil consumption, our research contributes to the broader field of nutrition and health, emphasizing the health benefits of incorporating extra virgin olive oil into dietary patterns.

## Fundings

The PREDIMED study was funded by NIH grants R01 HL118264, R01 DK102896 and R01 3R01 HL118264, the Spanish Ministry of Health (Instituto de Salud Carlos III, The PREDIMED Network grant RD 06/0045, 2006–2013, coordinated by M.A. Martínez-González; 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 (Projects CNIC-06/2007, CIBER 06/03, PI06-1326, PI07-0954, PI11/02505, SAF2009-12304, and AGL2010–22319-C03-03) and by the Generalitat Valenciana (PROMETEO17/2017 and PROMETEO 21/2021). Dr. Guasch-Ferré acknowledges that The Novo Nordisk Foundation Center for Basic Metabolic Research is supported by and unrestricted grant from the Novo Nordisk Foundation (grant no.

NNF18CC0034900). Dr. Salas-Salvadó gratefully acknowledges the financial support by ICREA under the ICREA Academia program.

## Abbreviations

95% CI	95% Confidence interval
AAMU	5-Acetamido-6-amino-3-methyl-uracil
ADP	Adenosine diphosphate
BMI	Body mass index
CE	Cholesterol ester
COO	Common olive oil
CV	Cross-validation
CVD	Cardiovascular diseases
DAG	Diacylglycerol
DMGV	Dimethylguanidino valeric acid
VOO	Virgin olive oil
FFQ	Food frequency questionnaires
HR	Hazard ratio
LC-MS	Liquid chromatography-tandem mass spectrometry
LPC	Lysophosphatidylcholine
LPE	Lysophosphatidylethanolamine
MAG	Monoacylglycerol
MEDAS	Mediterranean Diet Adherence Screener
MedDiet	Mediterranean Diet
MUFA	Monounsaturated fatty acid
PC	Phosphatidylcholine
PREDIMED	PREVención con Dieta MEDiterránea
PS	Phosphatidylserine
SM	Sphingomyelin
T2D	Type 2 diabetes
TAG	Triacylglycerol
TCA	Tricarboxylic acid cycle
TMAO	Trimethylamine N-oxide

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12933-023-02066-1>.

**Additional file1: Table S1.** Associations between plasma metabolites and total olive oil, extra virgin olive oil, and common olive oil consumption at baseline. **Table S2.** Hazard ratios (95% CIs) of type 2 diabetes incidence according to metabolomics profiles of energy-adjusted olive oil and its subtypes in the PREDIMED study groups stratified by intervention group. **Table S3.** Hazard ratios (95% CIs) of cardiovascular disease incidence according to metabolomics profiles of energy-adjusted olive oil and its subtypes in the PREDIMED study groups stratified by intervention group. **Figure S1.** Flowchart of participants and analysis. **Figure S2.** Volcano plot showing the associations between plasma metabolites and common olive oil consumption at baseline. **Figure S3.** Metabolite coefficients (mean and SD) selected ten times in the 10-cross validation of the continuous elastic regression for energy-adjusted common olive oil. **Figure S4.** Biplot of the principal component analysis using metabolites' coefficients derived by the elastic net continuous regression of each olive oil consumption approach. **Figure S5.** Correlation plot showing the correlation between each self-reported olive oil intake and olive oil metabolomics profiles and each food group consumption.

## Acknowledgements

The authors are grateful to the PREDIMED participants and personnel, as well as the affiliated primary care center staff, for their altruistic collaboration and excellent assistance.

## Author contributions

JGG, MGF, MR-C, CR, ET, MAM-G, JSS, FBH designed the research; DC, RE, MF, LSM, MAM-G, JSS coordinated the subject recruitment at the outpatient clinics and clinical data collection in Prevención con Dieta Mediterránea (PREDIMED);

CBC conducted the metabolomics data analysis; JGG conducted the statistical analysis; JGG, MGF, and JSS drafted the manuscript; JGG, and JSS, are the guarantors, and, as such, take responsibility for the integrity of the data and the accuracy of the data analysis; JGG, MAM-G, JSS, FBH had access to all the data in the study; JGG, NB, MF, J-PD-C, ET, CW, MAM-G, JSS, FBH interpreted the data; and all authors made critical revisions to the manuscript for key intellectual content and read and approved the final manuscript.

#### Availability of data and materials

The dataset generated and/or analyzed during the current study is not publicly available due to the lack of authorization from PREDIMED participants. Those wishing to access the PREDIMED trial data used in this study can request the corresponding author and it will then be passed on to members of the PREDIMED Steering Committee for deliberation.

#### Declarations

##### Ethics approval and consent to participate

The Research Ethics Committee of the Institut d'Investigacions Sanitàries Pere i Virgili approved the PREDIMED study protocol and all participants provided written informed consent.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare no conflict of interest in this article. The California Walnut Commission, Borges SA, and Morella Nuts SA donated nuts for the PREDIMED study. None of the funding sources played a role in the design, collection, analysis, or interpretation of the data or in the decision to submit the manuscript for publication.

##### Author details

<sup>1</sup>Universitat Rovira i Virgili, Departament de Bioquímica i Biotecnologia, Alimentació, Nutrició Desenvolupament i Salut Mental ANUT-DSM, Reus, Spain. <sup>2</sup>Institut d'Investigació Sanitària Pere Virgili (IISPV), Reus, Spain. <sup>3</sup>CIBER de Fisiopatologia de la Obesidad y Nutrición, Instituto de Salud Carlos III, Madrid, Spain. <sup>4</sup>Department of Preventive Medicine and Public Health, University of Navarra, IdiSNA, Pamplona, Spain. <sup>5</sup>Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA. <sup>6</sup>The Broad Institute of Harvard and MIT, Boston, MA, USA. <sup>7</sup>Department of Preventive Medicine, University of Valencia, Valencia, Spain. <sup>8</sup>Department of Internal Medicine, Institut d'Investigacions Biomèdiques August Pi Sunyer (IDIBAPS), Hospital Clinic, University of Barcelona, Barcelona, Spain. <sup>9</sup>Institut de Nutrició i Seguretat Alimentària (INSA-UB), Universitat de Barcelona, Barcelona, Spain. <sup>10</sup>Department of Endocrinology and Nutrition, Lipid Clinic, IDIBAPS, Hospital Clinic, University of Barcelona, Barcelona, Spain. <sup>11</sup>Cardiovascular and Nutrition Research Group, Institut de Recerca Hospital del Mar, Barcelona, Spain. <sup>12</sup>Bioaraba Health Research Institute, Osakidetza Basque Health Service, Araba University Hospital, University of the Basque Country UPV/EHU, 01009 Vitoria-Gasteiz, Spain. <sup>13</sup>Plataforma de Ensayos Clínicos, Instituto de Investigación Sanitaria Illes Balears (IdISBa), Hospital Universitario Son Espases, 07120 Palma, Spain. <sup>14</sup>Department of Family Medicine, Research Unit, Distrito Sanitario Atención Primaria Sevilla, Seville, Spain. <sup>15</sup>Polyphenol Research Group, Departament de Nutrició, Ciències de l'Alimentació i Gastronomia, Universitat de Barcelona (UB), Av. de Joan XXII, 27-31, 08028 Barcelona, Spain. <sup>16</sup>Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. <sup>17</sup>Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. <sup>18</sup>Department of Public Health, Section of Epidemiology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. <sup>19</sup>Faculty of Health and Medical Sciences, Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Øster Farimagsgade 5, 1014 Copenhagen, Denmark.

Received: 28 August 2023 Accepted: 14 November 2023

Published online: 13 December 2023

#### References

- Gaforio, Visioli, Alarcón-de-la-Lastra, Castañer, Delgado-Rodríguez, Fitó, et al. Virgin Olive Oil and Health: Summary of the III International Conference on Virgin Olive Oil and Health Consensus Report, JAEN (Spain) 2018. *Nutrients*. 2019 Sep 1;11(9):2039.
- Ros E. Olive oil and CVD: accruing evidence of a protective effect. *Br J Nutr*. 2012;108(11):1931–3.
- Martínez-González MA, Sayón-Orea C, Bullón-Vela V, Bes-Rastrollo M, Rodríguez-Artalejo F, Yusta-Boyo MJ, et al. Effect of olive oil consumption on cardiovascular disease, cancer, type 2 diabetes, and all-cause mortality: a systematic review and meta-analysis. *Clin Nutr*. 2022;41(12):2659–82.
- Schwingshackl L, Lampousi AM, Portillo MP, Romaguera D, Hoffmann G, Boeing H. Olive oil in the prevention and management of type 2 diabetes mellitus: a systematic review and meta-analysis of cohort studies and intervention trials. *Nutr Diabetes*. 2017;7(4):e262–e262.
- Guasch-Ferré M, Li Y, Willett WC, Sun Q, Sampson L, Salas-Salvadó J, et al. Consumption of olive oil and risk of total and cause-specific mortality among U.S. Adults *J Am Coll Cardiol*. 2022;79(2):101–12.
- Markellos C, Ourailidou ME, Gavriatopoulou M, Halvatsiotis P, Sergentanis TN, Psaltopoulou T. Olive oil intake and cancer risk: A systematic review and meta-analysis. *PLoS ONE*. 2022;17(1):e0261649.
- Schwingshackl L, Krause M, Schmucker C, Hoffmann G, Rucker G, Meerpohl JJ. Impact of different types of olive oil on cardiovascular risk factors: a systematic review and network meta-analysis. *Nutr Metab Cardiovasc Dis*. 2019;29(10):1030–9.
- Yubero-Serrano EM, Lopez-Moreno J, Gomez-Delgado F, Lopez-Miranda J. Extra virgin olive oil: more than a healthy fat. *Eur J Clin Nutr*. 2019;72(S1):8–17.
- Hernández Á, Fernández-Castillejo S, Farràs M, Catalán Ú, Subirana I, Montes R, et al. Olive oil polyphenols enhance high-density lipoprotein function in humans. *Arterioscler Thromb Vasc Biol*. 2014;34(9):2115–9.
- Fernández-Castillejo S, Pedret A, Catalán Ú, Valls R, Farràs M, Rubiò L, et al. Virgin olive oil phenolic compounds modulate the HDL lipidome in hypercholesterolaemic subjects: a lipidomic analysis of the VOHF study. *Mol Nutr Food Res*. 2021;65(9):2001192.
- Martínez-González MA, Corella D, Salas-Salvadó J, Ros E, Covas MI, Fiol M, et al. Cohort profile: design and methods of the PREDIMED study. *Int J Epidemiol*. 2012;41(2):377–85.
- Estruch R, Ros E, Salas-Salvadó J, Covas MI, Corella D, Arós F, 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(25):e34.
- Guasch-Ferré M, Ruiz-Canela M, Li J, Zheng Y, Bulló M, Wang DD, et al. Plasma acylcarnitines and risk of type 2 diabetes in a mediterranean population at high cardiovascular risk. *J Clin Endocrinol Metab*. 2019;104(5):1508–19.
- Ruiz-Canela M, Guasch-Ferré M, Toledo E, Clish CB, Razquin C, Liang L, 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.
- Guasch-Ferré M, Zheng Y, Ruiz-Canela M, Hruby A, Martínez-González MA, Clish CB, et al. Plasma acylcarnitines and risk of cardiovascular disease: effect of Mediterranean diet interventions. *Am J Clin Nutr*. 2016;103(6):1408–16.
- Ruiz-Canela M, Toledo E, Clish CB, Hruby A, Liang L, Salas-Salvadó J, et al. Plasma branched-chain amino acids and incident cardiovascular disease in the PREDIMED trial. *Clin Chem*. 2016;62(4):582–92.
- Willett W. *Nutritional epidemiology*. 3rd ed. New York: Oxford University Press; 2012.
- Fernández-Ballart JD, Piñol JL, Zazpe I, Corella D, Carrasco P, Toledo E, et al. Relative validity of a semi-quantitative food-frequency questionnaire in an elderly Mediterranean population of Spain. *Br J Nutr*. 2010;103(12):1808–16.
- Mataix J. *Tablas de composición de alimentos*. 4th ed. Granada, Spain: Universidad de Granada; 2003.
- Moreiras O, Carvajal A, Cabrera L. *Tablas de Composición de Alimentos [Food Composition Tables]*. 9th ed. Moreiras O, Carvajal A, Cabrera L, editors. Madrid, Spain: Ediciones Pirámide; 2005.
- Elosua R, Marrugat J, Molina L, Pons S, Pujol E. Validation of the minnesota leisure time physical activity questionnaire in spanish men. *Am J Epidemiol*. 1994;139(12):1197–209.

22. Hernández-Alonso P, Papandreou C, Bulló M, Ruiz-Canela M, Dennis C, Deik A, et al. Plasma metabolites associated with frequent red wine consumption: a metabolomics approach within the PREDIMED study. *Mol Nutr Food Res*. 2019;63(17):1900140.
23. O'Sullivan JF, Morningstar JE, Yang Q, Zheng B, Gao Y, Jeanfavre S, et al. Dimethylguanidino valeric acid is a marker of liver fat and predicts diabetes. *J Clin Investig*. 2017;127(12):4394–402.
24. Paynter NP, Balasubramanian R, Giulianini F, Wang DD, Tinker LF, Gopal S, et al. Metabolic predictors of incident coronary heart disease in women. *Circulation*. 2018;137(8):841–53.
25. Wang TJ, Larson MG, Vasani RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles and the risk of developing diabetes. *Nat Med*. 2011;17(4):448–53.
26. Stekhoven DJ, Buhlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28(1):112–8.
27. Gromski P, Xu Y, Kotze H, Correa E, Ellis D, Armitage E, et al. Influence of missing values substitutes on multivariate analysis of metabolomics data. *Metabolites*. 2014;4(2):433–52.
28. Wei R, Wang J, Su M, Jia E, Chen S, Chen T, et al. Missing value imputation approach for mass spectrometry-based metabolomics data. *Sci Rep*. 2018;8(1):663.
29. Blom G. Statistical estimates and transformed beta-variables. New York: Wiley; 1958.
30. Tripoli E, Giammanco M, Tabacchi G, Di Majo D, Giammanco S, La Guardia M. The phenolic compounds of olive oil: structure, biological activity and beneficial effects on human health. *Nutr Res Rev*. 2005;18(1):98–112.
31. Lozano-Castellón J, López-Yerena A, Rinaldi de Alvarenga JF, Romero del Castillo-Alba J, Vallverdú-Queralt A, Escribano-Ferrer E, et al. Health-promoting properties of oleocanthal and oleacein: two secoiridoids from extra-virgin olive oil. *Crit Rev Food Sci Nutr*. 2020;60(15):2532–48.
32. García-Aloy M, Hulshof PJM, Estruel-Amades S, Osté MCJ, Lankinen M, Geleijnse JM, et al. Biomarkers of food intake for nuts and vegetable oils: an extensive literature search. *Genes Nutr*. 2019;14(1):7.
33. Nikou T, Sakavitsi ME, Kalampokis E, Halabalaki M. Metabolism and bioavailability of olive bioactive constituents based on in vitro, in vivo and human studies. *Nutrients*. 2022;14(18):3773.
34. De la Torre R, Corella D, Castañer O, Martínez-González MA, Salas-Salvador J, Vila J, et al. Protective effect of homovanillyl alcohol on cardiovascular disease and total mortality: virgin olive oil, wine, and catechol-methyl-ation. *Am J Clin Nutr*. 2017;105(6):1297–304.
35. Almanza-Aguilera E, Davila-Cordova E, Guiñón-Fort D, Farràs M, Masala G, Santucci de Magistris M, et al. Correlation analysis between dietary intake of tyrosols and their food sources and urinary excretion of tyrosol and hydroxytyrosol in a European population. *Antioxidants*. 2023;12(3):715.
36. Soldevila-Domenech N, Boronat A, Mateus J, Diaz-Pellicer P, Matilla I, Pérez-Otero M, et al. Generation of the antioxidant hydroxytyrosol from tyrosol present in beer and red wine in a randomized clinical trial. *Nutrients*. 2019;11(9):2241.
37. Boronat A, Mateus J, Soldevila-Domenech N, Guerra M, Rodríguez-Morató J, Varon C, et al. Cardiovascular benefits of tyrosol and its endogenous conversion into hydroxytyrosol in humans. A randomized, controlled trial. *Free Radic Biol Med*. 2019;143:471–81.
38. Li J, Guasch-Ferré M, Chung W, Ruiz-Canela M, Toledo E, Corella D, et al. The Mediterranean diet, plasma metabolome, and cardiovascular disease risk. *Eur Heart J*. 2020;41(28):2645–56.
39. Meghwal M, Goswami TK. *Piper nigrum* and *Piperine*: an Update. *Phytother Res*. 2013;27(8):1121–30.
40. Talib WH, AlHur MJ, Al Naimat S, Ahmad RE, Al-Yasari AH, Al-Dalaeen A, et al. Anticancer effect of spices used in Mediterranean diet: preventive and therapeutic potentials. *Front Nutr*. 2022. <https://doi.org/10.3389/fnut.2022.905658>.
41. Bethlehem L, van Echten-Deckert G. Ectoin as novel anti-inflammatory and tissue protective lead compounds with special focus on inflammatory bowel disease and lung inflammation. *Pharmacol Res*. 2021;164:105389.
42. Bownik A, Stępniewska Z. Ectoin as a promising protective agent in humans and animals. *Arch Ind Hyg Toxicol*. 2016;67(4):260–5.
43. Cuozzo S, de LeBlanc AD, LeBlanc JG, Hoffmann N, Tortella GR. Streptomyces genus as a source of probiotics and its potential for its use in health. *Microbiol Res*. 2023;266:127248.
44. van Dorsten FA, Grün CH, van Velzen EJJ, Jacobs DM, Draijer R, van Duynhoven JPM. The metabolic fate of red wine and grape juice polyphenols in humans assessed by metabolomics. *Mol Nutr Food Res*. 2010;54(7):897–908.
45. Pallister T, Jackson MA, Martin TC, Zierer J, Jennings A, Mohney RP, et al. Hippurate as a metabolomic marker of gut microbiome diversity: modulation by diet and relationship to metabolic syndrome. *Sci Rep*. 2017;7(1):13670.
46. Tang WHW, Wang Z, Levison BS, Koeth RA, Britt EB, Fu X, et al. Intestinal Microbial Metabolism of Phosphatidylcholine and Cardiovascular Risk. *N Engl J Med*. 2013;368(17):1575–84.
47. Grizzell JA, Echeverria V. New Insights into the Mechanisms of action of cotinine and its distinctive effects from nicotine. *Neurochem Res*. 2015;40(10):2032–46.
48. Weimann A, Sabroe M, Poulsen HE. Measurement of caffeine and five of the major metabolites in urine by high-performance liquid chromatography/tandem mass spectrometry. *J Mass Spectrom*. 2005;40(3):307–16.
49. Wong P, Bachki A, Banerjee K, Leyland-Jones B. Identification of N1-methyl-2-pyridone-5-carboxamide and N1-methyl-4-pyridone-5-carboxamide as components in urine extracts of individuals consuming coffee. *J Pharm Biomed Anal*. 2002;30(3):773–80.
50. Jimenez-Lopez C, Carpena M, Lourenço-Lopes C, Gallardo-Gomez M, Lorenzo JM, Barba FJ, et al. Bioactive compounds and quality of extra virgin olive oil. *Foods*. 2020;9(8):1014.
51. An Y, Xu W, Li H, Lei H, Zhang L, Hao F, et al. High-fat diet induces dynamic metabolic alterations in multiple biological matrices of rats. *J Proteome Res*. 2013;12(8):3755–68.
52. Randle PJ, Garland PB, Hales CN, Newsholme EA. The glucose fatty-acid cycle its role in insulin sensitivity and the metabolic disturbances of *Diabetes mellitus*. *The Lancet*. 1963;281(7285):785–9.
53. Krzeminski R, Gorinstein S, Leontowicz H, Leontowicz M, Gralak M, Czerwinski J, et al. Effect of different olive oils on bile excretion in rats fed cholesterol-containing and cholesterol-free diets. *J Agric Food Chem*. 2003;51(19):5774–9.
54. Bae JS, Shin DH, Park PS, Choi BY, Kim MK, Shin MH, et al. The impact of serum uric acid level on arterial stiffness and carotid atherosclerosis: The Korean Multi-Rural Communities Cohort study. *Atherosclerosis*. 2013;231(1):145–51.
55. Lotito S, Frei B. Consumption of flavonoid-rich foods and increased plasma antioxidant capacity in humans: cause, consequence, or epiphenomenon? *Free Radic Biol Med*. 2006;41(12):1727–46.
56. Morze J, Wittenbecher C, Schwingshackl L, Danielewicz A, Rynkiewicz A, Hu FB, et al. Metabolomics and type 2 diabetes risk: an updated systematic review and meta-analysis of prospective cohort studies. *Diabetes Care*. 2022;45(4):1013–24.
57. McGranaghan P, Saxena A, Rubens M, Radenkovic J, Bach D, Schleußner L, et al. Predictive value of metabolomic biomarkers for cardiovascular disease risk: a systematic review and meta-analysis. *Biomarkers*. 2020;25(2):101–11.
58. Ruiz-Canela M, Guasch-Ferré M, Razquin C, Toledo E, Hernández-Alonso P, Clish CB, et al. Plasma acylcarnitines and risk of incident heart failure and atrial fibrillation: the Prevención con dieta mediterránea study. *Revista Española de Cardiología (English Edition)*. 2022;75(8):649–58.
59. Iob E, Steptoe A. Cardiovascular disease and hair cortisol: a novel biomarker of chronic stress. *Curr Cardiol Rep*. 2019;21(10):116.
60. Gall WE, Beebe K, Lawton KA, Adam KP, Mitchell MW, Nakhle PJ, et al.  $\alpha$ -Hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a nondiabetic population. *PLoS ONE*. 2010;5(5):e10883.
61. Guasch-Ferré M, Hu FB, Martínez-González MA, Fitó M, Bulló M, Estruch R, et al. Olive oil intake and risk of cardiovascular disease and mortality in the PREDIMED Study. *BMC Med*. 2014;12(1):78.
62. Urpi-Sarda M, Casas R, Chiva-Blanch G, Romero-Mamani ES, Valderas-Martínez P, Arranz S, et al. Virgin olive oil and nuts as key foods of the Mediterranean diet effects on inflammatory biomarkers related to atherosclerosis. *Pharmacol Res*. 2012;65(6):577–83.
63. Gimeno E, de la Torre-Carbot K, Lamuela-Raventós RM, Castellote AI, Fitó M, de la Torre R, et al. Changes in the phenolic content of low density lipoprotein after olive oil consumption in men. A randomized crossover controlled trial. *Br J Nutr*. 2007;98(6):1243–50.

64. Bullo M, Lamuela-Raventos R, Salas-Salvado J. Mediterranean diet and oxidation: nuts and olive oil as important sources of fat and antioxidants. *Curr Top Med Chem*. 2011;11(14):1797–810.
65. Salas-Salvadó J, Bulló M, Babio N, Martínez-González MÁ, Ibarrola-Jurado N, Basora J, et al. Reduction in the incidence of type 2 diabetes with the mediterranean diet. *Diabet Care*. 2011;34(1):14–9.
66. Salas-Salvadó J, Bulló M, Estruch R, Ros E, Covas MI, Ibarrola-Jurado N, et al. Prevention of diabetes with mediterranean diets. *Ann Intern Med*. 2014;160(1):1–10.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

