Metabolism 170 (2025) 156195

Contents lists available at ScienceDirect

Metabolism

journal homepage: www.journals.elsevier.com/metabolism

A multi-metabolite signature robustly predicts long-term mortality in the PREDIMED trial and several US cohorts

Gonzalo Fernández-Duval [a,b,*], Cristina Razquin [a,c], Fenglei Wang [d], Huan Yun [e], Jie Hu [e], Marta Guasch-Ferré [d,f,g], Kathryn Rexrode [h,i], Raji Balasubramanian [h,i], Jesús García-Gavilán [c,j,k], Miguel Ruiz-Canela [a,c], Clary B. Clish [l], Dolores Corella [c,m], Enrique Gómez-Gracia [n], Miquel Fiol [c,o], Ramón Estruch [c,p], José Lapetra [c,q], Montse Fitó [c,r], Luis Serra-Majem [c,s], Emilio Ros [c,p], Liming Liang [e,t], Courtney Dennis [l], Eva M. Asensio [c,m], Olga Castañer [c,r], Francisco J. Planes [b,u,v], Jordi Salas-Salvadó [c,j,k], Frank B. Hu [d,e,w], Estefanía Toledo [a,c], Miguel A. Martínez-González [a,c,d]

[a] Department of Preventive Medicine and Public Health, Navarra Health Research Institute (IdiSNA), University of Navarra, Pamplona, Spain
[b] Institute of Data Science and Artificial Intelligence (DATAI), University of Navarra, Pamplona, Spain
[c] Consorcio CIBER, Fisiopatología de la Obesidad y Nutrición (CIBERObn), Instituto de Salud Carlos III (ISCIII), Madrid, Spain
[d] Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[e] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[f] Department of Public Health, Section of Epidemiology, University of Copenhagen, Copenhagen, Denmark
[g] Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark
[h] Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
[i] Division of Women's Health, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
[j] Universitat Rovira i Virgili, Departament de Bioquímica i Biotecnologia, Unitat de Nutrició Humana, Reus, Spain
[k] Institut d'Investigació Sanitària Pere Virgili (IISPV), Reus, Spain
[l] Metabolomics Platform, Broad Institute of MIT and Harvard, Cambridge, MA, USA
[m] Department of Preventive Medicine and Public Health, University of Valencia, Valencia, Spain
[n] Department of Preventive Medicine, University of Malaga, Instituto de Investigación Biomédica de Málaga (IBIMA), Málaga, Spain
[o] Platform for Clinical Trials, Instituto de Investigación Sanitaria Illes Balears (IdISBa), Hospital Universitario Son Espases, Palma de Mallorca, Spain
[p] Institut d'Investigacions Biomèdiques August Pi Sunyer (IDIBAPS), Barcelona, Spain
[q] Department of Family Medicine, Research Unity, Distrito Sanitario Atención Primaria Sevilla, Sevilla, Spain
[r] Hospital del Mar Research Institute, Barcelona, Spain
[s] Nutrition Research Group, Research Institute of Biomedical and Health Sciences (IUIBS), University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain
[t] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[u] Tecnun School of Engineering, University of Navarra, San Sebastián, Spain
[v] Biomedical Engineering Center, University of Navarra, Pamplona, Spain
[w] Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

ARTICLE INFO

Keywords:
Plasma metabolites
Metabolomic
Metabolomic signature
Biomarkers
All-cause mortality

ABSTRACT

Metabolome-based biomarkers contribute to identify mechanisms of disease and to a better understanding of overall mortality. In a long-term follow-up subsample ($n = 1878$) of the PREDIMED trial, among 337 candidate baseline plasma metabolites repeatedly assessed at baseline and after 1 year, 38 plasma metabolites were identified as predictors of all-cause mortality. Gamma-amino-butyric acid (GABA), homoarginine, serine, creatine, 1-methylnicotinamide and a set of sphingomyelins, plasmalogens, phosphatidylethanolamines and cholesterol esters were inversely associated with all-cause mortality, whereas plasma dimethylguanidino valeric acid (DMGV), choline, short and long-chain acylcarnitines, 4-acetamidobutanoate, pseudouridine, 7-methylguanine, N6-acetyllysine, phenylacetylglutamine and creatinine were associated with higher mortality. The multi-metabolite signature created as a linear combination of these selected metabolites, also showed a strong association with all-cause mortality using plasma samples collected at 1-year follow-up in PREDIMED. This

* Corresponding author at: Department of Preventive Medicine and Public Health, University of Navarra. Irunlarrea 1, 31008, Pamplona, Navarra, Spain.
E-mail address: ghfernandezd@unav.es (G. Fernández-Duval).

https://doi.org/10.1016/j.metabol.2025.156195
Received 24 November 2024; Accepted 15 March 2025
Available online 17 March 2025

association was subsequently confirmed in 4 independent American cohorts, validating the signature as a consistent predictor of all-cause mortality across diverse populations.

1. Introduction

Metabolomics has proven to be a reliable way of understanding biological pathways leading to chronic disease, aging process, and mortality [1–3]. Metabolomics-based biomarkers have shown adequate predictability of multi-disease outcomes and they contribute to the identification of high-risk individuals [4]. Recent findings suggest a high potential for translational research and clinical precision medicine through the identification of metabolites related to premature mortality across animal models and human subjects [5]. Additionally, several plasma metabolites have been associated with all-cause mortality and longevity, including acylcarnitines, some amino acids, phospholipids, and purine/pyrimidines [6–8]. However, whether these associations are independent, replicable or if they provide useful clinical information beyond conventional risk factors remains elusive. In the context of a worldwide trend of increased overall life expectancy during the second half of the twentieth century and the first decades of the twenty first century, this type of research has also become essential to identify and understand the metabolic alterations underlying the aging process and representing additional independent predictors of overall mortality [9]. In addition, the association between a metabolomic signature and the risk of all-cause mortality has not been addressed in Mediterranean regions with long life expectancies likely associated with adherence to a healthy traditional Mediterranean dietary pattern.

Previous studies have shown evidence supporting the importance of metabolomic profiles for identifying the risk of chronic disease and specific causes of death. In particular, in a subsample of participants followed-up from 2003 to 2010 in the PREDIMED ("Prevención con Dieta Mediterránea") randomized trial, metabolomics techniques were used to identify a metabolomic signature of the Mediterranean diet (MedDiet) adherence screener (MEDAS). The identified signature (a linear combination of 67 plasma metabolites), reflecting closer adherence to the MedDiet, was strongly and robustly associated with a lower risk of cardiovascular disease (CVD) during the trial period (2003−2010). That inverse association of the MedDiet with CVD was also replicated in independent large American cohorts: the Nurses' Health Study I (NHS-I) and II (NHS-II), and the Health Professionals Follow-Up Study (HPFS). These cohorts consisted of initially healthy subjects who were at lower cardiovascular risk compared to the participants in the PREDIMED trial [10]. Interestingly, the MedDiet, an established high-quality dietary pattern, has also been related to healthy aging [11].

Beyond those findings, and given that an expanded mortality follow-up is now available for participants in the PREDIMED trial up to 2019, we sought to identify a metabolomic signature to predict all-cause mortality and may provide further molecular insights into the aging processes.

In this context, our aim was to identify a plasma metabolite profile independently associated with all-cause mortality, after controlling for phenotypic data in a long-term follow-up of participants in the PREDIMED trial, and then to replicate these findings in four independent US cohorts.

2. Methods

2.1. PREDIMED trial

The primary assessment was conducted in PREDIMED (PREvención con DIeta MEDiterránea), a well-known multicenter, randomized, primary cardiovascular prevention trial with Mediterranean diets. Table 1 shows baseline characteristics of participants and Supplementary Table 1 describes characteristics of PREDIMED participants at 1-year follow-up. From 2003 to 2010, this large Spanish trial was conducted in 11 recruiting centers with the participation of 7447 men (aged 55 to 80 years) and women (aged 60 to 80 years), at high cardiovascular risk but initially free of CVD [12]. They were randomized 1:1:1 into three intervention groups: MedDiet supplemented with extra-virgin olive oil (EVOO), MedDiet supplemented with nuts, and control group (advice on a low-fat diet) [13].

2.2. NHS-I, NHS-II and HPFS cohorts

The NHS-I began in 1976, including 121,700 U.S. female registered nurses aged 30–55 years; the NHS-II was initiated in 1989, enrolling 116,429 female registered nurses aged 25–42 years; and the HPFS was established in 1986, recruiting 51,525 male health professionals aged 40–75 years. Supplementary Tables 2 and 3 describe baseline characteristics of participants. These cohorts have been described in detail elsewhere [14]. Blood samples were collected from 32,826 NHS-I participants between 1989 and 1990 [15], 29,611 NHS-II participants between 1996 and 1999 [15], and 18,225 HPFS participants between 1993 and 1995 [16].

2.3. Women's health initiative

The Women's Health Initiative (WHI) is a long-term nationwide study that focuses on strategies for the prevention and control of the most common causes of morbidity and mortality among post-menopausal women [17]. During 1993 to 1998 in 40 clinical centers in the U.S., a total of 161,808 women aged 50–79 years were enrolled into either an observational study (OS) (n = 93,676) or the hormone replacement therapy trials (HT) or dietary medication trial (n = 68,132). Supplementary Table 4 describes baseline characteristics of participants. The WHI-HT trials included a randomized, placebo-controlled trial of conjugated equine estrogens plus medroxyprogesterone acetate (n = 16,608), and a randomized, double-blind, placebo-controlled disease prevention trial of 0.625 mg/day of conjugated equine estrogens (n = 10,739) [18].

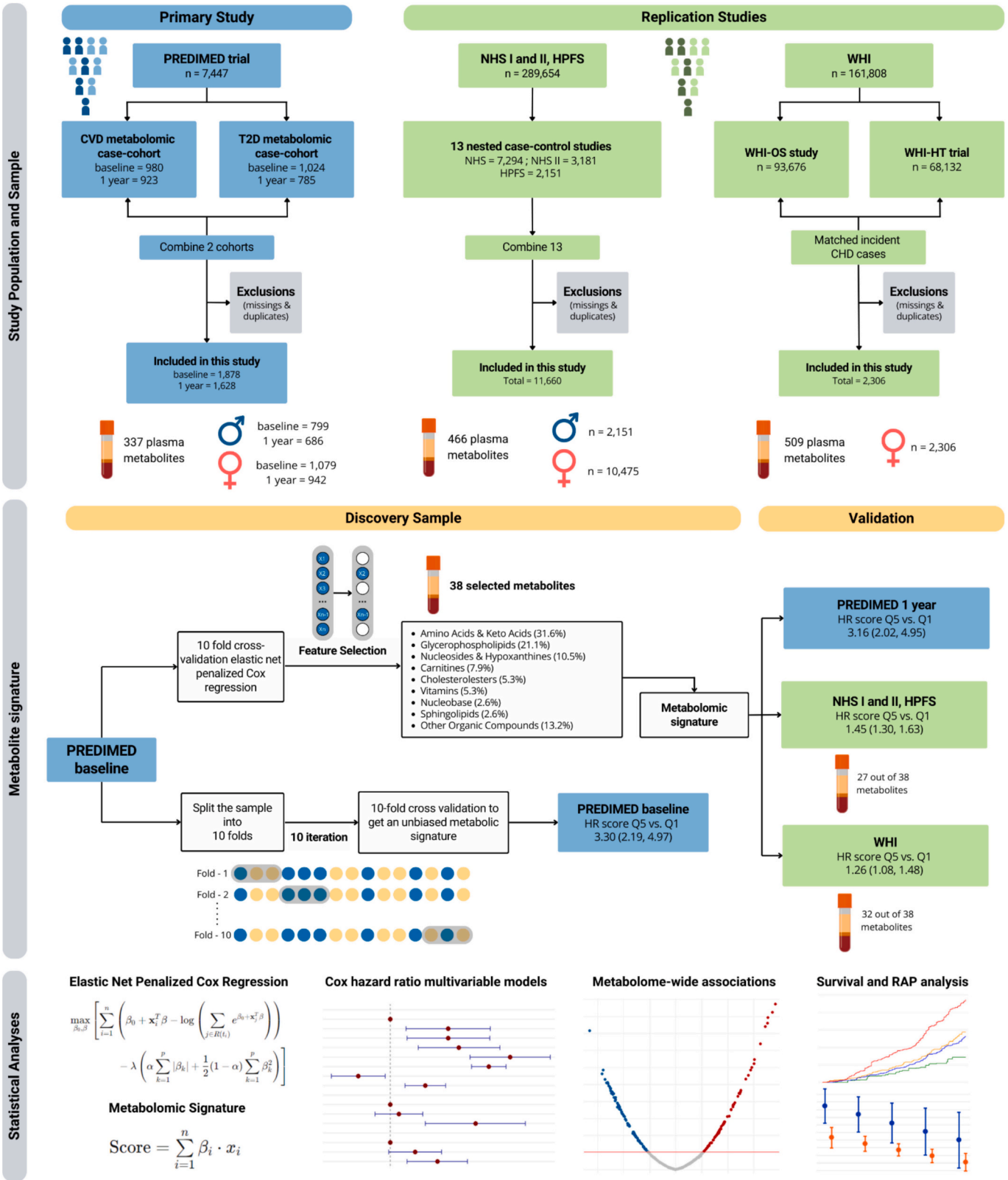
2.4. Study participants

Two case-cohort studies [19], nested within the PREDIMED trial, were designed for metabolite profiling [20]: PREDIMED-CVD (including 229 incident cardiovascular disease cases and a random subcohort of 788 participants with 37 overlapping cases) and PREDIMED-T2D (including 251 incident type 2 diabetes cases and a random subcohort of 694 participants with 53 overlapping cases). Additional participants with available oral glucose tolerance tests were also included for metabolomics analyses (n = 132). After excluding duplicates and participants with >20 % missing values in metabolites, 1878 participants with metabolomic information were available in PREDIMED. Among them, 457 all-cause deaths were observed during a median follow-up period of 12.2 years (up to June 30, 2019). As an internal replication, we used metabolomics data from PREDIMED after 1 year of intervention, for 1628 participants who also had repeated measurements of diet, covariates and plasma metabolomics. The protocol was approved by the Institutional Review Boards at all PREDIMED study locations, and all participants provided written informed consent.

For the current study, NHS-I, NHS-II and HPFS baseline date was set as the respective blood draw date for each participant. Participants from 13 nested case-control studies on metabolomics were included (Supplementary Table 3). All participants in these studies were free of the

Table 1
Baseline characteristics of PREDIMED participants included in the analyses, by quintiles of the multi-metabolite score, and by sex. Mean (standard deviation) values for continuous variables. Percentage values by quintiles of the metabolomic score and by sex. ANOVA test for continuous values by quintiles of the metabolomic score, *t*-test for continuous variables by sex and Chi-squared test for proportion (percentage) variables.

	Quintiles of the leave-one-fold baseline score						Sex			Overall
	Q1	Q2	Q3	Q4	Q5	*P* value	Male	Female	P value	
N	376	376	376	375	375		799	1079		1878
Mean score (SD)	−0.66 (0.21)	−0.28 (0.08)	−0.02 (0.07)	0.24 (0.08)	0.71 (0.28)	<0.001	0.17 (0.48)	−0.13 (0.47)	<0.001	0.00 (0.49)
Mean age (SD)	64 (5)	66 (5)	66 (6)	68 (6)	70 (6)	<0.001	66 (6)	67 (5)	<0.001	67 (6)
Female sex %	80.3	64.9	58.0	46.4	37.6	<0.001	–	–	–	57.5
All-cause mortality %	8.2	16.8	20.7	26.7	49.3	<0.001	30.9	19.5	<0.001	24.3
Cancer mortality %	2.9	5.3	6.9	7.7	12.8	<0.001	9.5	5.4	<0.001	7.1
CVD mortality %	2.9	4.3	6.4	10.1	21.1	<0.001	11.6	7.0	<0.001	8.9
Mortality from non-cancer non-CVD causes %	2.4	7.2	7.5	8.8	15.5	<0.001	9.8	7.1	0.050	8.3
Type 2 diabetes %	16.0	21.0	29.3	32.0	47.7	<0.001	33.8	25.8	<0.001	29.2
Dyslipidemia %	82.2	83.5	74.2	73.9	70.1	<0.001	71.6	80.6	<0.001	76.8
Hypertension %	87.5	85.9	87.8	86.1	89.3	0.625	84.5	89.4	0.002	87.3
Family history of premature coronary heart disease %	28.2	30.1	23.9	23.5	18.4	0.002	16.8	30.8	<0.001	24.8
Smoking %						<0.001			<0.001	
Never	71.8	66.0	56.4	56.3	46.1		23.5	85.8		59.3
Former	11.2	18.3	26.9	25.1	36.8		49.1	6.9		24.8
Current	17.0	15.7	16.8	18.7	17.1		27.4	7.3		15.9
Primary education or lower %	93.4	91.0	92.0	92.8	94.4	0.435	88.9	95.6	<0.001	92.7
Mean baseline glucose (SD), mg/dL	105.7 (22.1)	108.9 (27.7)	114.2 (33.5)	117.1 (35.5)	126.3 (42.1)	<0.001	118.3 (34.0)	111.6 (33.1)	<0.001	114.4 (33.6)
Mean alcohol intake (SD), g/d	6.2 (11.3)	8.2 (12.9)	9.3 (14.7)	11.4 (17.1)	13.1 (19.6)	<0.001	17.5 (20.0)	3.8 (6.9)	<0.001	9.6 (15.6)
Mean BMI (SD), kg/m2	29.7 (3.3)	29.7 (3.6)	30.1 (3.6)	29.9 (3.9)	30.1 (3.5)	0.366	29.4 (3.2)	30.3 (3.8)	<0.001	29.9 (3.6)
Mean waist-to-height ratio (SD)	0.62 (0.06)	0.62 (0.06)	0.63 (0.06)	0.63 (0.06)	0.64 (0.06)	<0.001	0.62 (0.06)	0.63 (0.07)	<0.001	0.63 (0.06)
Mean total energy intake level (SD), kcal/d	2223(580)	2308 (585)	2345 (632)	2321 (584)	2383 (622)	0.006	2508 (637)	2174 (533)	<0.001	2316 (603)
Mean leisure-time physical activity level (SD), MET-min/d	221 (187)	245 (226)	265 (282)	259 (258)	232 (223)	0.068	329 (289)	182 (167)	<0.001	245 (238)
Mean Mediterranean Diet Score	8.7 (2)	8.6 (2)	8.6 (2)	8.6 (2)	8.5 (2)	0.767	8.7 (2)	8.6 (2)	0.125	8.6 (2)
Randomized Group %						0.356			0.274	
Olive Oil	31.7	37.0	33.0	34.1	36.0		34.3	34.4		34.3
Nuts	38.8	34.3	34.3	34.7	29.9		36.2	33.1		34.4
Control (low fat)	29.5	28.7	32.7	31.2	34.1		29.5	32.5		31.3



(caption on next page)

Fig. 1. Scheme of the study development. Using a combination of two PREDIMED case-cohorts as the main data, the association between baseline plasma metabolites and all-cause mortality was addressed through metabolome-wide association approaches and also feature selection by an elastic net penalized Cox regression model, which make it possible to create a metabolomic signature (score) using these selected metabolites, separating training sets from testing sets, with special care to avoid any overfitting. Further analysis was implemented using an unbiased metabolic signature, evaluating the prediction of all-cause mortality with Cox hazard multivariable models. Additionally, an assessment of this metabolomic signature in relation to premature mortality was conducted using rate advancement period (RAP) and survival analyses. Internal validation of a plasma 38-multimetabolite score was conducted using another sample of PREDIMED (PREDIMED specimens collected after 1 year of follow-up instead of baseline). External validations were implemented in American cohorts, using 13 combined studies of Nurses' Health Studies I and II and Health Professionals Follow-Up Study (HPFS), and a case-control study of coronary heart disease nested in 2 combined cohorts from the Women's Health Initiative.
For Elastic Net Penalized Cox Regression, α and λ are fitting parameters to control the L1-norm and L2-norm penalization of the model, $R(t_i)$: risk set at time t_i. For Metabolomic Signature, x_i: abundance of each metabolite; β_i: weight of each metabolite to the metabolomic score, n: number of selected metabolites.
PREDIMED: PREvención con DIeta MEDiterránea; NHS-I: Nurses´ Health Study I; NHS-II: Nurses´ Health Study II; HPFS: Health Professionals Follow-Up Study.

disease under study at the time of blood collection but were selected if they developed the endpoint after blood collection or were selected as an age-matched control. We excluded participants with a history of CVD or cancer at baseline or lost to follow-up after blood collection. A total of 11,660 participants remained in the prospective mortality risk analyses. The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required.

The WHI study included 1153 incident CHD cases and an equal number of matched controls (matching factors: age, self-reported race, hysterectomy status, and enrollment groups). All participants were free of known CVD and cancer at baseline. The research protocol was approved by the Fred Hutchinson Cancer Research Center Institutional Review Board and the Institutional Review Board of Mass General Brigham/Brigham and Women's Hospital. Informed consent was obtained from all participants.

2.5. Metabolite profiling

The fasting plasma metabolomics profiling in the PREDIMED, NHS-I, NHS-II and HPFS cohorts was conducted using three liquid chromatography tandem mass spectrometry (LC–MS) techniques (HILIC-pos, HILIC-neg, and C8-pos) at the Broad Institute of MIT and Harvard (Cambridge, MA) during 2015–2021 [21,22]. To account for any temporal drift in instrument performance over time and between batches, pooled plasma reference samples (prepared by combining small aliquots from the study samples) were analyzed every 20 participant samples. Additionally, quality control (QC) samples, to which the laboratory was blinded, were randomly distributed among the study samples for profiling. For the WHI, plasma samples were collected using EDTA-containing tubes and then stored at −70 °C before analysis. Metabolomics analyses of WHI samples were performed using the same LC-MS methods used in PREDIMED during 2013–2015. Pooled plasma QC samples were analyzed after intervals of 20 WHI samples to ensure repeatability and to scale data between batches. Detailed metabolomic profiling have been described for the PREDIMED [22], NHS/NHSII/HPFS [22,23] and WHI [22,24].

In PREDIMED, 337 labeled metabolites from the HILIC-pos (cationic polar metabolites) and C8 (lipids) platforms were selected for primary analyses after quality filtering and standardization. In an ancillary analysis, we also added other 41 candidate metabolites measured at baseline in PREDIMED using the HILIC-negative method. In NHS-I, NHS-II and HPFS, a total of 466 metabolites of known identity were annotated. Metabolites for which quality control replicates intraclass correlation coefficient was <0.3 ($n = 10$) or detection rate was <70 % ($n = 171$) were excluded. The final number of metabolites considered in the primary analysis was 243. In WHI, after quality control, 509 known metabolites were used in current analyses.

Criteria for exclusion in the PREDIMED and WHI were metabolites with at least 20 % of total missing values. Remaining metabolites were imputed using half the minimum detected of each metabolite value. Finally, they were standardized using Blom's method (inverse-normal transformation) prior to statistical analysis. Specifically, for the NHS-I, NHS-II and HPFS cohort, metabolites with a missing proportion higher than 30 % were excluded. Metabolite data were log-transformed if its distribution was highly skewed (defined as absolute skewness >2). Then all metabolites were converted to z-scores within each sub-study and imputed missing data using the random forest imputation approach [25].

2.6. Measurements of covariates

In PREDIMED, participants self-reported lifestyle factors. Medical records, risk factors and family history of diseases were collected during the first screening visit. At baseline and during annual visits, MedDiet adherence was measured by a validated 14-item questionnaire [26] in a face-to-face interview with a registered dietitian, and anthropometric traits were measured by trained study personnel. Physical activity was collected with validated questionnaires. Hypertension was defined as blood pressure > 140/90 mmHg or treatment with antihypertensive drugs. Dyslipidemia was defined as LDL-C levels >4.14 mmol/l (>160 mg/dl) or treatment with hypolipidemic agents; HDL-C concentrations <1.29 mmol/l (50 mg/dl) for women or < 1.03 mmol/l (40 mg/dl) for men independently of lipid-lowering therapy. Diagnosis of type-2 diabetes was based on at least one of the following criteria: current treatment with insulin or oral hypoglycemic drugs; fasting glucose >126 mg/dl (fasting is defined as no caloric intake at least for 8 h); casual glucose >200 mg/dl with polyuria, polydipsia, or unexplained weight loss; or glucose >200 mg/dl in two measurements after an oral glucose tolerance test. Due to the high proportion of participants with type 2 diabetes or prediabetes in the PREDIMED trial, baseline glucose levels were used as a covariate for multivariable adjustment.

In NHS-I, NHS-II and HPFS, information on body weight, smoking status, physical activity, multivitamin use, race, diabetes, hypertension, hypercholesterolemia, and the use of antihypertensive medication and lipid-lowering medication was obtained through self-reported questionnaires prior to blood collection. Body mass index (BMI) was calculated using the height reported at cohort baseline and the body weight reported before the blood draw. Information on age and fasting status was obtained via questionnaires completed at blood collection. Total calories, alcohol intake, and the Alternate Healthy Eating Index (AHEI, a measure of overall diet quality ranging from 0 to 100, not including alcohol), were calculated from the last available semiquantitative food-frequency questionnaire (FFQ) before the blood draw. The validity and reproducibility of the FFQs have been reported elsewhere [27].

In the WHI study, information on age, race/ethnicity, lifestyle factors and other covariates, including smoking status, alcohol consumption, physical activity, education, female-specific variables (e.g., hysterectomy, menopausal status, hormone use), health status (diabetes, hypertension, dyslipidemia), medication use (hormone, multivitamin, aspirin) were collected at the study baseline. Diet was assessed from the year 3 annual follow-up food frequency questionnaire [28]. Body weight, height, waist circumference, and hip circumference were measured by trained staff at baseline. Medical history was collected semiannually during the intervention main phase and annually during extension studies.

2.7. Ascertainment of death

The primary outcome was all-cause death, separated by cause: CVD, cancer or other. Each category was added to consider all-cause mortality. In PREDIMED, this information was identified from different sources: annual review of medical records, direct repeated follow-up of participants, and a yearly consultation of the National Death Index. Information on fatal events was submitted to the Clinical End-Point Committee (whose members were blinded to the intervention-group assignment) for validation. Materially complete follow-up for all-cause mortality was available until June 30, 2019 from the National Death Index under an agreement with the National Institute of Statistics.

Deaths in the NHS-I, NHS-II and HPFS were identified from state vital statistics records, the National Death Index, and reports from next of kin or the postal authorities. The follow-up for mortality in these cohorts is over 98 % complete using these methods. The cause of death was determined by physician's review of medical records, autopsy reports, or death certificates. We used the International Classification of Diseases, Eighth Revision (ICD-8) in NHS and ICD-9 in HPFS, which were the ICD systems used at the time the cohorts began.

In WHI, death information was linked with the National Death Index. Adjudication of diseases was performed by trained physicians first at the Clinical Centers, and then under the auspices of the Clinical Coordinating Center. Specific cause of death was assessed by physician adjudicators or by relatives' reports.

2.8. Association of individual metabolites with all-cause mortality

Out of 337 available metabolites in PREDIMED, the individual association between each metabolite (per +1 SD increase) and all-cause mortality was studied using a Cox regression for each individual metabolite and estimating the association with all-cause mortality and specific death causes (cancer, CVD and other causes). Also, analyses of all-cause mortality were stratified by sex. Corrections for multiple testing were addressed using the Benjamini & Hochberg procedure under a < 0.05 false discovery rate (FDR) criteria.

2.9. Metabolomic profile score

To better address potential bias and variance, while taking advantage of both features of the Ridge and Lasso penalized regressions, the machine learning method used to select metabolites as consistent predictors of all cause-mortality was a 10-fold cross validation elastic net Cox regression [29,30]. As their β coefficients were used as weights, for the validation samples, each individual score was calculated as the sum of the measured metabolite values (previously inverse normally transformed) weighted by their corresponding β coefficient. In the training sample, to avoid overfitting [31], the score was calculated using a leave-one-fold-out (LOFO) approach. For this aim we divided the sample into 10 folds, and then applied a 10-iteration 10-fold cross-validation elastic net Cox regression. We calculated the score for each of the ten left-out folds separately using the other 9 folds as training set and applied the coefficients independently of the derivation set to the left-out fold in each iteration (90 % of the sample was considered in each iteration to calculate the remaining 10 % score values).

2.10. Metabolomic score and all-cause mortality

Cox proportional hazard models were run to analyze the association between the metabolite score calculated and all-cause mortality in the main sample (baseline data of PREDIMED) and in replication samples. For replication samples, we used a 38-metabolite score derived from the elastic net Cox regression using the baseline PREDIMED sample as derivation set. Cox models were run with three different levels of adjustment: crude (bi-variable, only the score), adjusted for age and sex, and multivariable-adjusted, further adjusted for baseline glucose levels, dyslipidemia (yes or no), hypertension (yes or no), family history of premature CHD (yes or no), smoking (never, former and current), educational level (low or high), alcohol intake (<5 g/d, 5 to 15 g/d and > 15 g/d), BMI (<30, 30–35, >35 kg/m^2), waist-to-height ratio (3 categories, <0.6, 0.6 to <0.75 and > 0.75), total energy intake (per 1-SD), physical activity (per 1-SD), score of adherence to MedDiet (0 to 14 points) and randomized diet group (olive oil, nuts or control diet fat). For better interpretation, the score was tested in quintiles and also per +1 SD.

Finally, as a second metabolite selection step, an additional, more parsimonious, version of the score was generated using stepwise (backward and forward) Cox regression on the 38 previously selected metabolites. The score was then calculated using LOFO coefficients, including only the 12 newly selected metabolites.

2.11. Ten-year follow-up association with all-cause mortality

The association between all-cause mortality and the LOFO-metabolomic score, survival after 10 years of follow-up in the PREDIMED cohort, was analyzed using logistic regression and ROC curve analysis. The comparison of multivariable models with and without the score was performed using DeLong's ROC test and the Hosmer-Lemeshow goodness of fit for calibration.

2.12. Mortality assessment, survival analysis, and estimation of the rate advancement period

A graphical procedure for survival analysis was conducted using inverse probability weighting to estimate the probability of the exposure observed for a particular participant, the inverse of this predicted probability was used as the weighting factor in a survival model. For a better graphical understanding of the results, score quintiles 2 and 3 (materially equivalent) were merged in a single category. Also, a rate advancement period [32] (RAP) estimation was performed, stratified by specific baseline age groups (in groups of 5 years, from 55 to 59 years to 75–79 years). To test if there was an interaction between the exposure (score) and baseline age [33], Martingale's residuals from a Cox hazard ratio were analyzed assessing the association of the score (per +1 SD) with all-cause mortality. For further comparison, different RAPs were estimated for participants in two ways: comparing the sample according to whether participants were above or below the median of the multi-metabolite score, and comparing the fifth versus first quintile of the multi-metabolite score.

2.13. Subgroup analyses

As an alternative to sensitivity analyses, and to avoid misinterpretation of the data, we performed a series of subgroup analyses. First, we use conventional Cox models and competing risk models to analyze the association between the score and each particular cause of death (CVD, cancer and other causes), representing graphically incidence curves, using the score as a continuous variable (per +1 SD) and also comparing (HRs) the upper and the lowest quintile. Also, we ran Cox regression models after stratifying by sex and categories of baseline age. Finally, to discard major influences of the main risk factors (type 2 diabetes, smoking, hypertension, dyslipidemia and obesity) a new model was designed to analyze the association of the score and a categorical variable counting the presence of these risk factors (3 categories, 1, 2, 3 o more) was included as a covariate.

2.14. Replication into independent cohorts

As an internal validation of the ability of our metabolomic score to predict the risk of mortality, the same methodology for assessing the association of the score with mortality was replicated with metabolomics data of PREDIMED after 1 year of intervention instead of using

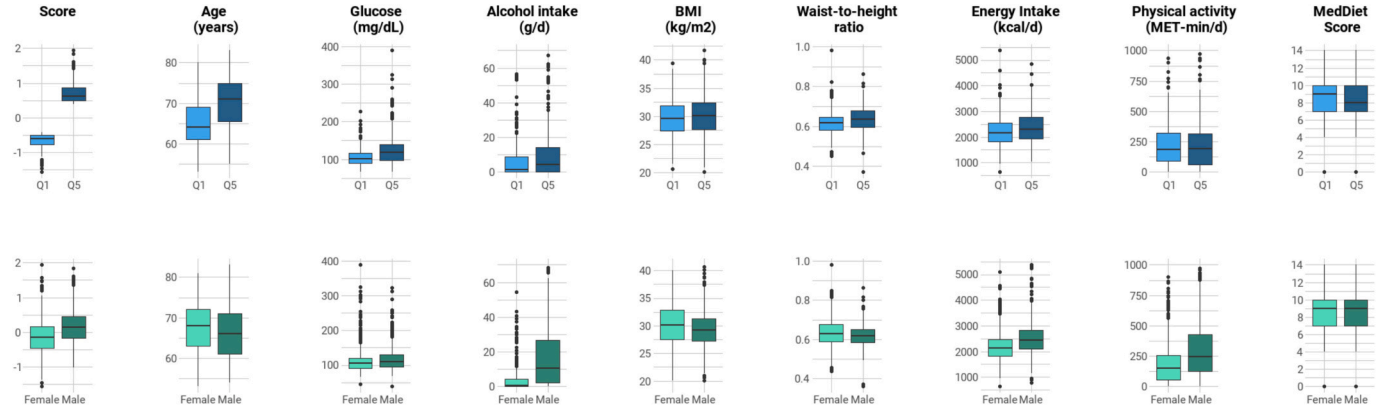


Fig. 2. Boxplots of baseline continuous variables in each extreme quintile of the metabolomic score (first row) and stratified by sex (second row).

the baseline metabolomics data. External validations were conducted in the NHS-I, NHS-II, HPFS and WHI cohorts. Particularly, 27 metabolites out of the 38 selected for the score were available in NHS-I, NHS-II and HPFS (two were not available: choline and xanthosine; and nine had over 30 % missing values: *N*-acetylputrescine, hydroxycotinine, trimethylbenzene, DMGV, CAR18:1 OH, 1-methyladenosine, CAR18:1, PC-plasmalogen 36:1 and guanine). In the case of WHI, a total of 32 of the initial 38 metabolites were available. Missing metabolites were hydroxycotinine, guanine, GABA, DMGV, PC 40:9 and alpha-aminoisobutyric acid. As part of the sensitivity analysis, we reconstructed the PREDIMED multi-metabolite score using only the 24 metabolites available in all US cohorts, allowing to assess the performance of the score when restricted to the available metabolites in each of these external validation cohorts. As a sensitivity analysis, different baseline scoring tests were performed in PREDIMED with the metabolites discarded in the other cohorts.

2.15. Analysis implementation

All analyses were implemented using R version 4.4.1, using the main following libraries: 'glmnet 4.1–8' for Elastic Net regressions, 'survival 3.6–4' for survival analysis, 'caret 6.0–94' and 'pROC 1.18.5' for logistic regressions, 'cmprsk 2.2–12' and 'tidycmprsk 1.1.0' for competing risk analysis and 'tidyverse 2.0.0' features for data transformation and graphics.

3. Results

3.1. Study participants

Our primary study sample was made of a subsample of baseline PREDIMED data including 1878 participants (mean age = 67 years, SD = 5 years) with available baseline fasting plasma metabolomic information (337 available metabolites). For 1628 of these participants, repeated plasma measurements of the 337 metabolites after 1 year were also available and they were used for internal validation (Fig. 1). Predominantly, participants had similar baseline characteristics as those in the whole roster ($N = 7447$) of the PREDIMED trial [12] (Table 1, Fig. 2 and Supplementary Table 1). As expected, given the selection of a high-risk population for this trial, long-term (2003 to 2019) mortality was substantial, and higher for men (30.9 %) than for women (19.5 %).

For external validation, harmonized data from a total of 11,660 participants in 13 combined studies nested in the NHS-I, NHS-II, and HPFS were used, as well as 2306 participants from the combined female cohorts of the WHI. In contrast to PREDIMED, participants from the Harvard cohorts (NHS-I, NHS-II and HPFS) were younger (mean = 54 years, SD = 9 years) and had lower baseline prevalence of obesity, diabetes, dyslipidemia and hypertension (Supplementary Table 2 and Supplementary Table 3), while women participating in the WHI study were age similar to female participants of PREDIMED (Supplementary Table 4).

3.2. Identification of individual metabolites associated with all-cause mortality

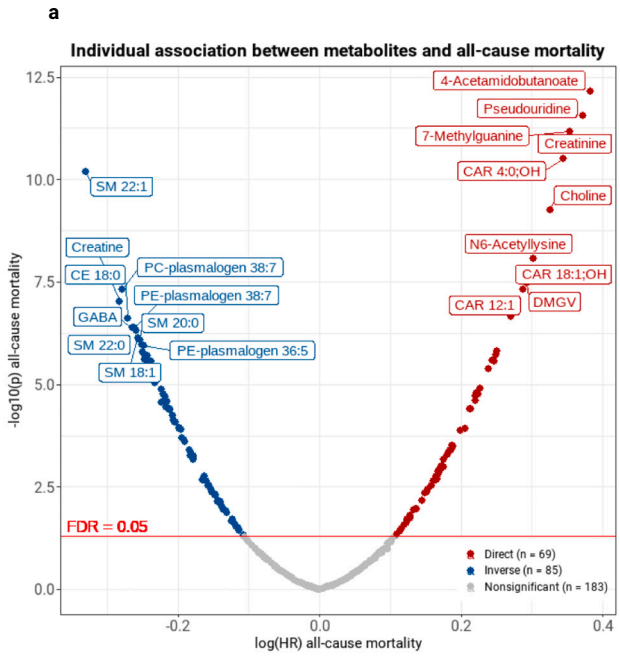
To assess the association of each individual plasma metabolite with all-cause mortality, a metabolome-wide association analysis was conducted. We also assessed specific causes of mortality (cancer, CVD and other causes).

When we assessed each individual metabolite in PREDIMED, a total of 154 metabolites were associated with all-cause mortality (after correcting for multiple testing), out of which 85 showed inverse associations and 69 showed direct associations with all-cause mortality, after applying a false discovery rate (FDR) < 0.05 (Fig. 3.a. and Supplementary Table 5). Plasma metabolites with strong inverse associations with all-cause mortality included SM 22:1, PC-plasmalogen 38:7, several other cholesterol esters, glycerophospholipids, creatine, gamma-aminobutyric acid (GABA), and other amino acids. On the other hand, direct associations with all-cause mortality were observed for several acylcarnitines, dimethylguanidino valeric acid (DMGV), nucleosides, hypoxanthines, pseudouridine, 7-methylguanine, and other amino and keto acids as in 4-acetamidobutanoate and creatinine. Results remained consistent for both women and men, with a Pearson r for beta coefficients between men and women = 0.72 (Fig. 3.b. and Supplementary Table 6).

Regarding specific causes of mortality, only 13 metabolites were associated with cancer mortality (7 inverse associations and 6 direct associations) (Fig. 3.b. and Supplementary Table 7). For CVD-specific mortality, 105 metabolites showed significant associations (50 inverse associations and 55 direct associations) (Fig. 3.c. and Supplementary Table 8). Finally, 54 metabolites were found to be associated with mortality from non-cancer non-CVD causes (28 inverse associations and 26 direct associations) (Fig. 3.d. and Supplementary Table 9).

3.3. Metabolomic signature for all-cause mortality used in replications

Using a 10-fold cross-validation elastic net penalized Cox regression, a group of 38 baseline plasma metabolites were identified and selected to develop a plasma multi-metabolomic signature (score) able to predict all-cause mortality (Supplementary Table 10 and Supplementary Fig. 1). This 38-multi-metabolite score was calculated as the weighted sum of the selected metabolites with weights equal to the elastic net Cox regression coefficients (Supplementary Table 11). High Pearson correlations were found between metabolites of glycerophospholipid, nucleoside and hypoxanthine families (Supplementary Fig. 2). This linear combination composed of 38 plasma metabolites with their weights



Inverse associations: lower mortality (Top 20)		Positive associations: higher mortality (Top 20)	
Metabolite	**HR (95% CI)**	**Metabolite**	**HR (95% CI)**
SM 22:1	0.72 (0.66, 0.79)	4-Acetamidobutanoate	1.47 (1.33, 1.61)
Creatine	0.75 (0.69, 0.83)	Pseudouridine	1.45 (1.32, 1.59)
PC-plasmalogen 38:7	0.76 (0.69, 0.83)	7-Methylguanine	1.42 (1.30, 1.56)
CE 18:0	0.76 (0.69, 0.84)	Creatinine	1.42 (1.30, 1.56)
GABA**	0.77 (0.70, 0.84)	CAR 4:0;OH	1.41 (1.29, 1.55)
PE-plasmalogen 38:7ˣ	0.77 (0.70, 0.84)	Choline*	1.38 (1.26, 1.52)
SM 22:0ˣ	0.77 (0.70, 0.85)	N6-Acetyllysine	1.35 (1.23, 1.49)
SM 18:1ˣ	0.77 (0.71, 0.85)	CAR 18:1;OH*	1.35 (1.23, 1.48)
SM 20:0ˣ	0.78 (0.71, 0.85)	DMGV***	1.34 (1.22, 1.47)
PE-plasmalogen 36:5	0.78 (0.71, 0.85)	CAR 12:1ˣ	1.33 (1.21, 1.46)
Homoarginine	0.78 (0.71, 0.85)	N-Acetylputrescine*	1.32 (1.20, 1.45)
TG 54:10ˣ	0.78 (0.71, 0.85)	1-Methyladenosine*	1.31 (1.19, 1.44)
Serine	0.78 (0.71, 0.86)	CAR 14:1ˣ	1.28 (1.17, 1.41)
SM 24:0ˣ	0.78 (0.71, 0.86)	1-Methylguanosineˣ	1.28 (1.17, 1.41)
CE 20:4	0.78 (0.72, 0.86)	PE 34:2	1.28 (1.16, 1.40)
CE 22:6ˣ	0.79 (0.72, 0.86)	CAR 18:1*	1.28 (1.16, 1.40)
PC-plasmalogenB 36:5ˣ	0.79 (0.72, 0.86)	N-Carbamoyl-beta-alanineˣ	1.27 (1.16, 1.39)
CE 18:2ˣ	0.79 (0.72, 0.87)	Phenylacetylglutamine	1.26 (1.14, 1.38)
1-Methylnicotinamide	0.79 (0.72, 0.87)	CAR 12:0ˣ	1.25 (1.14, 1.37)
PC-plasmalogen 38:6ˣ	0.79 (0.73, 0.87)	CAR 10:0ˣ	1.25 (1.14, 1.37)

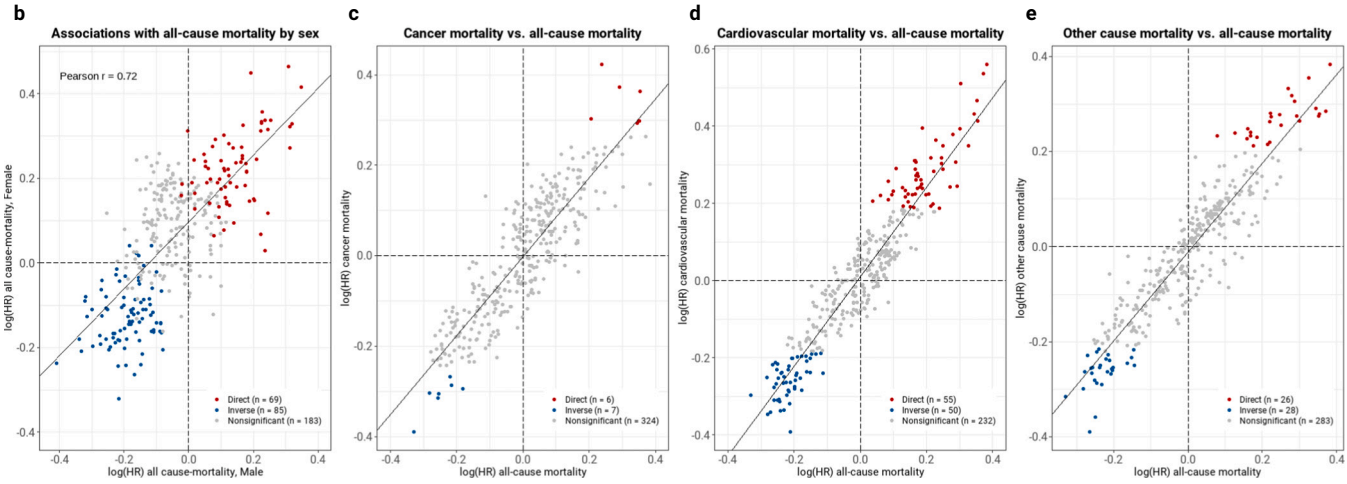


Fig. 3. Associations between each metabolite and mortality. a. Volcano plot for the individual associations between each metabolite and long-term all-cause mortality; The top 40 associations (20 inverse and 20 direct hazard ratios) of individual metabolites with mortality and their respective HRs (95 % CI) per +1 SD increment are shown in the upper left graph and the upper right table. **b. Scatter plot** to compare the results of the beta coefficients [log(HR)] of Cox models for each metabolite by sex. Direct and inverse significant associations for both male and female (FDR < 0.05) with individual metabolite and all-cause mortality are represented with red and blue, respectively **c. d. & e.** Scatters plot for log(HR) of all-cause mortality vs. log(HR) of each cause of death (cancer, CVD and other causes). Direct and inverse significant associations for each cause of death (FDR < 0.05) with individual metabolite and all-cause mortality are represented in red (direct) and blue (inverse). All *p*-values were corrected using the Benjamini & Hochberg FDR < 0.05 correction; In the upper right table, we indicate the non-included or non-available metabolites: ˣnot selected for the 38-multi-metabolomic signature associated with all-cause mortality; *not available in NHS-I, NHS-II and HPFS; **not available in WHI, ***not available in NHS-I, NHS-II, HPFS and WHI. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(negative or positive) derived from a single elastic net penalized Cox regression applied to baseline PREDIMED was used thereafter for replications in PREDIMED after 1 year (but not in the baseline assessment in PREDIMED, see in 3.4), and also in external independent cohorts (using those metabolites that were available in each cohort, out of the 38 identified metabolites).

3.4. Leave-one-fold out score (LOFO) in PREDIMED at baseline

To avoid overfitting in the original PREDIMED sample at baseline, the multi-metabolite score was initially derived by employing a 10-fold cross-validation method. The sample was divided into 10 folds and we performed 10 iterations of an elastic net Cox regression. For each iteration, 9 folds were used to derive the coefficients, while the remaining

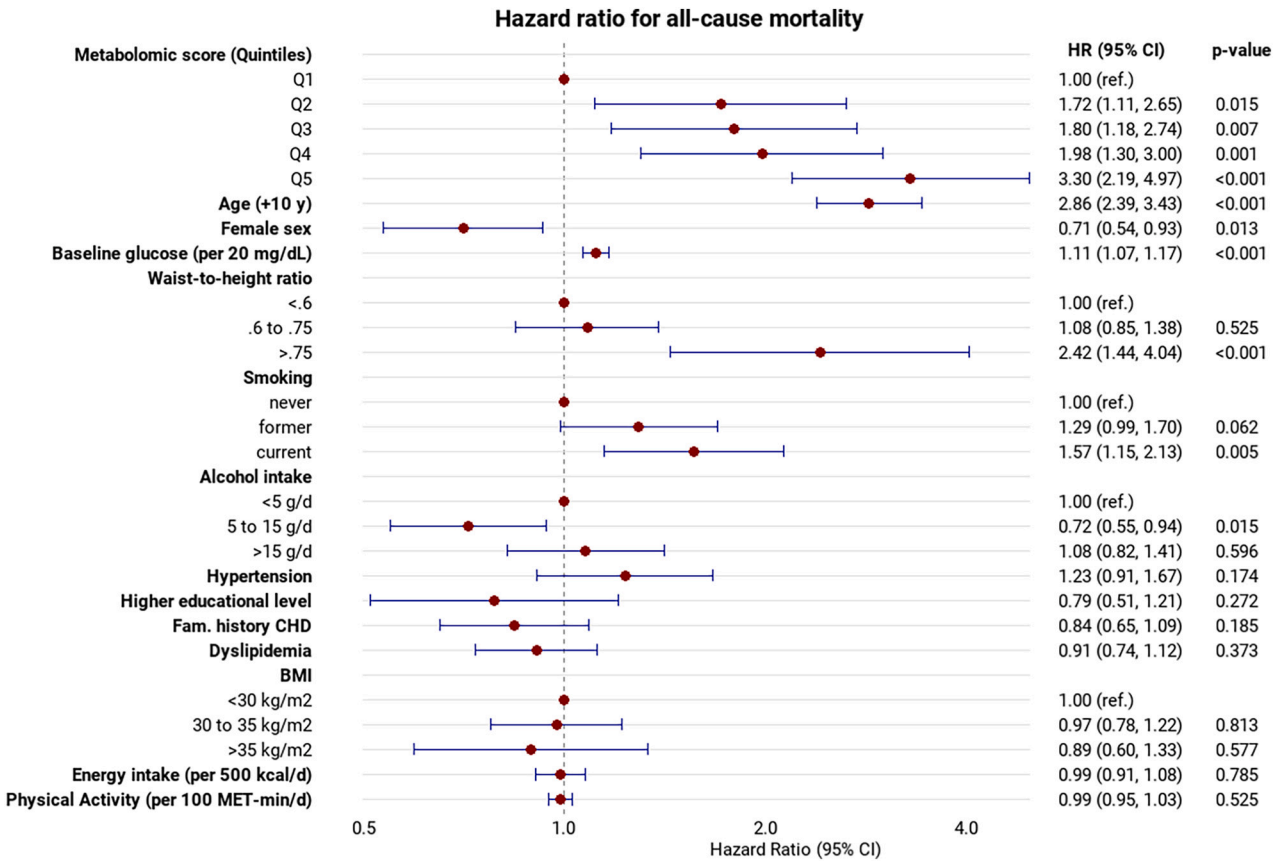


Fig. 4. Multivariable-adjusted model showing independent associations between baseline variables, including the plasma LOFO-multi-metabolite score, and all-cause mortality, adjusted for the conventional risk factors shown in the table and graph. Hazard ratio (HR) for quintiles of the LOFO-multi-metabolite score, adjusted for all the variables shown in the graph: age, sex, baseline glucose level (per 20 mg/dL), waist-to-height ratio (3 categories: <0.6, 0.6 to <0.75 and ≥0.75), smoking (3 categories: never, former and current), alcohol intake (3 categories: <5 g/d, 5 to 15 g/d and ≥15 g/d), hypertension [HT] at baseline (yes or no), educational level (low or high), family history of premature coronary heart disease [CHD] (yes or no), dyslipidemia diagnosed at baseline (yes or no), body mass index [BMI] (3 categories:<30, 30–35, ≥ 35 kg/m^2), total energy intake (per 500 kcal/d) and leisure-time physical activity (per 100 MET-min/d). In addition, the shown model was also adjusted for baseline adherence to the Mediterranean diet adherence screener (0–14 points) and for the randomized group (3 categories: olive oil, nuts and low fat). HR: Hazard ratio; CI: Confidence interval. CHD: Coronary heart disease.

fold was left out for testing (leave-one-fold-out, LOFO). The score for each of the ten folds, which were left out, was calculated by applying the coefficients obtained in the remaining 90 % of the sample (the other 9 folds) from which that fold was excluded. This procedure ensured that the LOFO score was independently calculated from the derivation set in each of the 10 left out folds. The number of metabolites included in each of the ten different folds ranged between 31 and 47 metabolites (Supplementary Table 12). A comparison of each individual coefficient between the 38-multimetabolite score and the LOFO-multi-metabolite score was made (Supplementary Table 13) and the results were consistent (Supplementary Fig. 3).

The prediction of all-cause mortality using this LOFO-multi-metabolite score was first assessed by quintiles. The results showed a significant association with long-term mortality for each of the four upper quintiles as compared to the lowest quintile. Specifically, the multivariable-adjusted hazard ratio (HR) for the fifth versus the first quintile was 3.30 (95 % Confidence interval [CI]: 2.19, 4.97). Additionally, in the continuous analysis, the HR per +1 SD was 1.45 (95 % CI: 1.31, 1.61) (Fig. 4). C-concordance indexes for the multivariable model were satisfactory and similar for the score in quintiles (0.761) and for the continuous analysis per +1 SD (0.761).

Consistent results were found when doing a new metabolite selection that included HILIC-neg metabolites. This process led to the identification of seven new metabolites and the exclusion of three from the previous selection, resulting in a score with a total of 42 metabolites (Supplementary Tables 14 and 15, and Supplementary Fig. 4), Similarly, consistent results were found with a parsimonious 12-metabolite score obtained with stepwise Cox regression (Supplementary Table 16) and when comparing specific causes of mortality (Supplementary Table 17). Also, when the sample was stratified by the number of risk factors, the results were consistent (Supplementary Table 18).

The magnitude of the association between the LOFO-multimetabolite score was not homogenous when stratifying by sex (Supplementary Fig. 5), with a stronger association in women than in men. Subgroup analysis showed no significant effect modification by the main covariables included in the model on the effect of the score on all-cause mortality, except for categorized baseline age, showing a stronger effect of the multi-metabolite signature among younger than among older subjects, with $p = 0.004$ for interaction (Supplementary Table 19). Finally, using mortality at 10-year follow-up, we compared in ROC curves, the area under the curve (AUC) for the prediction of death by conventional risk factor with or without the LOFO-multi-metabolite score, and we found a significant improvement ($p = 0.01$ for the comparison of the AUC) when we added the multi-metabolite score to the conventional predictors of all-cause death (Supplementary Fig. 6). Also, the calibration and goodness of fit of the model were correct according to the Hosmer-Lemeshow method (Supplementary Fig. 7).

Table 2
Associations between the multi-metabolite score and all-cause mortality in the PREDIMED trial and in replication cohorts.

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5		per + 1 SD	
	HR (95 % CI)	HR (95 % CI)	HR (95 % CI)	HR (95 % CI)	HR (95 % CI)	P for trend	HR (95 % CI)	P value
PREDIMED baseline (LOFO score)[a)]								
Cases/participants	31/376	63/376	78/376	100/375	185/375		457/1878	
Crude model	1.00 (ref.)	2.07 (1.35, 3.18)	2.63 (1.74, 3.99)	3.53 (2.36, 5.28)	8.04 (5.49, 11.76)	<0.001	1.94 (1.78, 2.12)	<0.001
Age & sex adjusted	1.00 (ref.)	1.69 (1.10, 2.60)	1.93 (1.26, 2.93)	2.19 (1.45, 3.31)	3.92 (2.62, 5.87)	<0.001	1.54 (1.39, 1.70)	<0.001
Multivariable adjusted	1.00 (ref.)	1.72 (1.11, 2.65)	1.80 (1.18, 2.74)	1.98 (1.30, 3.00)	3.30 (2.19, 4.97)	<0.001	1.45 (1.31, 1.61)	<0.001
PREDIMED 1 year (38-metabolite score)[a)]								
Cases/participants	26/326	39/326	56/326	89/325	154/325		364/1628	
Crude model	1.00 (ref.)	1.51 (0.92, 2.49)	2.23 (1.40, 3.55)	3.81 (2.46, 5.89)	7.57 (5.00, 11.48)	<0.001	2.21 (2.01, 2.44)	<0.001
Age & sex adjusted	1.00 (ref.)	1.28 (0.78, 2.10)	1.71 (1.07, 2.74)	2.40 (1.53, 3.77)	3.78 (2.44, 5.87)	<0.001	1.79 (1.60, 2.00)	<0.001
Multivariable adjusted	1.00 (ref.)	1.21 (0.74, 1.99)	1.69 (1.06, 2.70)	2.07 (1.31, 3.27)	3.16 (2.02, 4.95)	<0.001	1.69 (1.51, 1.90)	<0.001
NHSI/NHSII/HPFS (27-metabolite score)[b)]								
Cases/participants	814/2332	744/2332	784/2332	885/2332	1154/2332		4381/11660	
Crude model	1.00 (ref.)	1.09 (0.98, 1.20)	1.21 (1.09, 1.34)	1.41 (1.27, 1.56)	1.77 (1.59, 1.98)	<0.001	1.28 (1.23, 1.33)	<0.001
Age & sex adjusted	1.00 (ref.)	1.11 (1.00, 1.23)	1.18 (1.06, 1.31)	1.42 (1.28, 1.58)	1.61 (1.44, 1.79)	<0.001	1.23 (1.19, 1.28)	<0.001
Multivariable adjusted	1.00 (ref.)	1.10 (0.99, 1.21)	1.11 (1.00, 1.23)	1.32 (1.18, 1.47)	1.45 (1.30, 1.63)	<0.001	1.19 (1.14, 1.23)	<0.001
WHI (32-metabolite score)[c)]								
Cases/participants		341/461	347/461	344/462	377/460		1686/2306	
Crude model	1.00 (ref.)	1.08 (0.93, 1.26)	1.02 (0.88, 1.19)	1.19 (1.02, 1.39)	1.28 (1.10, 1.49)	<0.001	1.09 (1.04, 1.15)	<0.001
Age adjusted	1.00 (ref.)	1.07 (0.92, 1.25)	0.94 (0.81, 1.10)	1.08 (0.92, 1.26)	1.16 (1.00, 1.35)	0.075	1.06 (1.01, 1.11)	0.027
Multivariable adjusted	1.00 (ref.)	1.13 (0.97, 1.33)	0.96 (0.82, 1.13)	1.11 (0.94, 1.31)	1.26 (1.08, 1.48)	0.012	1.07 (1.02, 1.13)	0.007

HR: Hazard ratio; CI: Confidence interval; LOFO: Leave-one-fold out; PREDIMED: PREvención con DIeta MEDiterránea; NHS-I: Nurses´ Health Study I; NHS-II: Nurses´ Health Study II; HPFS: Health Professionals Follow-Up Study; CHD: Coronary heart disease.

[a] *PREDIMED baseline and 1 year*. Multivariable model adjusted for age, sex, glucose, waist-to-height ratio (3 categories: <0.6, 0.6 to <0.75 and ≥0.75), smoking (3 categories: never, former and current), alcohol intake (3 categories: <5 g/d, 5 to 15 g/d and ≥15 g/d), baseline hypertension (yes or no), educational level (low or high), family history of premature coronary heart disease (yes or no), baseline dyslipidemia (yes or no), body mass index (3 categories:<30, 30–35, ≥ 35 kg/m^2), total energy intake, physical activity level, Mediterranean diet score (0–14) and randomized group (3 categories: olive oil, nuts and low fat).

[b] *NHSI/NHSII/HPFS*: Crude model was stratified by case/control status, endpoint, and cohort. Multivariable model adjusted for age, sex, race, fasting status, multivitamin use, smoking, body mass index, alcohol intake, baseline diabetes, baseline hypertension, dyslipidemia, anti-hypertensive medication, lipid-lowering medication, family history of coronary heart disease, modified alternative healthy eating index (no alcohol), and total energy intake.

[c] *WHI:* Crude model was stratified by CHD case-control status, WHI sub-studies and intervention arms. Multivariate model adjusted for age, WHI sub-studies and intervention arms, CHD case-control status, race/ethnicity, fasting status, education level, smoking (3 categories: never, former and current), alcohol intake (3 categories: <5 g/d, 5 to 15 g/d and > 15 g/d), hormone therapy use, physical activity level (per 1-SD), diabetes (yes or no), hypertension (yes or no), dyslipidemia (yes or no), total energy intake (per 1-SD), waist-to-height ratio (per 1-SD), body mass index (3 categories:<25, 25–30, >30).

3.5. Validation of the plasma metabolomic signature

Different assessments were conducted to replicate the observed association between the identified 38-multi-metabolite score and mortality. As an internal validation, the association of the 38-multi-metabolite score with subsequent all-cause death was replicated in PREDIMED participants using plasma metabolomic profiles measured after 1 year of follow-up (1628 participants). A highly consistent association of the 38-multi-metabolite score with subsequent mortality (>1 year follow-up) was found, both when comparing extreme quintiles (HR = 3.16; 95 % CI = 2.02, 4.95; C-index = 0.756) and per +1 SD score (HR = 1.69; 95 % CI = 1.51, 1.90; C-index = 0.766), after adjusting for potential confounders (Table 2).

Furthermore, several external replications were conducted. In the combined sets of participants from the NHS-I, NHS-II, and HPFS, 27 of the 38 metabolites were available. The weighted score built with these 27 metabolites was significantly associated with higher mortality, when comparing extreme quintiles (HR = 1.45; 95 % CI = 1.30, 1.63; C-index = 0.722) and per +1 SD score (HR = 1.19; 95 % CI = 1.14, 1.23; C-index = 0.722), after adjusting for relevant covariables (Table 2). Similarly, significant results were observed in the WHI CHD case-control study, where the score was based on 32 of the 38 original metabolites, when comparing extreme quintiles (HR = 1.26; 95 % CI = 1.08, 1.48; C-index = 0.847) and per +1 SD (HR = 1.07; 95 % CI = 1.02, 1.13; C-index = 0.847), after adjusting for relevant covariables (Table 2). When we formally assessed the heterogeneity of the 3 HRs (PREDIMED, NHS-I and II and HPFS, and WHI) by one SD in the metabolomic score, they showed statistically significant heterogeneity according to the Q statistics (Q = 29.1, df = 2, $p < 0.001$; I squared = 0.93).

As a sensitivity analysis, using LOFO coefficients for each individual,

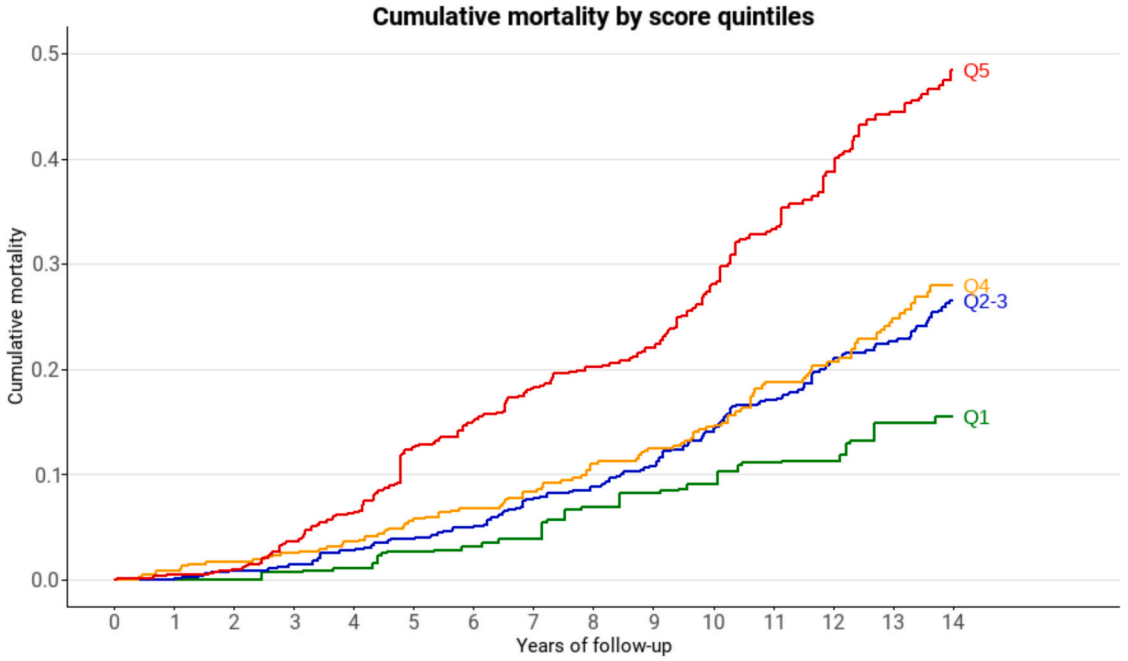


Fig. 5. Cumulative all-cause mortality by quintiles of the baseline LOFO-multi-metabolite score. Long-term mortality rates in PREDIMED by quintiles (Q1 to Q5) of the baseline multi-metabolomic score, adjusted for potential confounders (the same variables shown in the footnote of Table 2) using inverse probability weighting (IPW). Quintiles 2 and 3 were merged because they mostly overlapped.

we reconstructed the baseline multi-metabolite score in PREDIMED using only the 24 metabolites available in the all combined US cohorts. The multivariable-adjusted HR was 2.37 (95 % CI = 1.63, 3.45) for the fifth versus the lowest score quintile and 1.37 (95 % CI = 1.24, 1.52) per +1 SD score increase. We also built a parsimonious 12 metabolite score (score 12) using only those baseline metabolites selected by a stepwise Cox model. We obtained similar results in PREDIMED with this score 12. The multivariable-adjusted HR was 3.11 (95 % CI = 2.09, 4.65) for the fifth versus the lowest score 12 quintile and 1.50 (95 % CI = 1.35, 1.66) per +1 SD of the score 12 increase. Finally, we made an alternative version of these 12 metabolites score excluding 6 metabolites which were unavailable in the combined US cohorts: the multivariable-adjusted HR in PREDIMED was 1.84 (95 % CI = 1.31, 2.58) for the fifth versus the lowest score quintile and 1.29 (95 % CI = 1.16, 1.43) per +1 SD score increase (Supplementary Fig. 8).

3.6. Mortality assessment

When participants were grouped in quintiles by the identified baseline LOFO-multi-metabolite score, they showed a monotonic pattern of increased cumulative mortality across successive quintiles of this score, after adjusting for confounders using inverse probability weighting (Fig. 5). Almost identical curves for the second and third quintiles were found, so they were merged.

The association of the LOFO-multi-metabolite score with all-cause death showed effect modification by baseline age, suggesting a higher impact on earlier rather than on later mortality. In fact, an effect modification by baseline age [34] due to the interaction was suggested after analyzing the Martingale residuals of the original Cox hazard results (Supplementary Fig. 9). Consequently, specific comparisons for the multi-metabolite score were conducted after stratifying by baseline age categories with 3 different criteria for stratification (Supplementary Table 20). Using 70 years of baseline age as the cut-off point, the results were significant in both strata. For participants younger than 70 years, the multivariable-adjusted HR comparing extreme quintiles was 5.34 (95 % CI: 3.04 to 9.38), and for participants 70 years or older, it was 3.89 (2.16to 7.01). Alternative cut-off points provided analogous results (Supplementary Table 20).

To further develop this concept, within each stratum of baseline age, we estimated the rate advancement period (RAP) by quintiles of the multi-metabolite score and by values above or below the median. Results showed a stronger impact (in terms of the RAP) on estimates of premature mortality for younger baseline age groups than for older baseline categories of age (Supplementary Fig. 10).

4. Discussion & conclusions

In an expanded follow-up for a large subset of participants at high cardiovascular risk in the PREDIMED trial, we were able to identify a linear combination of plasma metabolites strongly and robustly associated with all-cause death, particularly in women and younger subjects. This association with mortality was stronger for the multi-metabolite score than for chronological age or any conventional risk factor. These results were replicated for the linear combination of 38 metabolites after 1-year follow-up in PREDIMED. Though weaker associations were found, these results were replicated with statistically significant associations for similar combinations of metabolites (some of them were unavailable) in several external and independent cohorts conducted in the US, mostly composed of participants with a considerably lower burden of cardiovascular risk factors.

The individual plasma metabolites most strongly associated with reduced mortality included sphingomyelins (SM 22:1, SM 22:0, SM 18:1, SM 20:0, SM 24:0), plasmalogens (PC-plasmalogenB 36:5, PC-plasmalogen 38:7, PC-plasmalogen 38:6, PE-plasmalogen 38:7, PE-plasmalogen 36:5) and cholesterol esters (18 CE:0, 20 CE:4, 22 CE:6 and 18 CE:2). Other molecules also strongly inversely associated with long-term mortality were GABA, homoarginine, serine, creatine and 1-methylnicotinamide. In addition, a triglyceride composed of polyunsaturated fatty acids (TG 54:10) showed an inverse association.

Dysfunction of endothelial cells is a key mechanism in atherogenesis and CVD development and sphingolipid metabolism disorders exert this effect on endothelial cells [35]. Sphingomyelin (SM), a ubiquitous component of cells, is involved in various cellular activities, including cell division, proliferation, and autophagy. It also helps maintain a

balance between pro-inflammatory and anti-inflammatory lipids, thus regulating the immune system. Moreover, tumor-necrosis factor (TNF)-alpha signaling usually leads to SM hydrolysis to ceramide catalyzed by sphingomyelinases. Inflammation increases ceramides by up-regulating the activity of sphingomyelinases or sphingomyelin phosphodiesterases [36]. Consequently, reduced plasma levels of some SMs and increased ceramide levels may reflect a pro-inflammatory status. Also, elevated levels of sphingomyelins containing longer-chain fatty acids (SM-20, 22 and 24) were associated with lower mortality risk [37]. In addition, some studies have reported that ceramides are a major contributing factor to insulin resistance and type 2 diabetes [38]. In fact, higher ceramide levels were associated with CVD risk in the PREDIMED trial [39] whereas high SM levels were associated with better physical performance in the ARIC cohort [40]. Also, when SM levels decrease and ceramide levels increase, there is a pathological progression toward liver steatosis and fibrosis formation in the context of fatty liver disease, which is very common in patients with diabetes, hypertension and a high burden of CVD risk factors, such as the PREDIMED cohort. Under these and other mechanistic assumptions, it does not seem surprising that higher plasma levels of some SMs were included in the set of metabolites associated with longer survival.

GABA, an important inhibitory neurotransmitter in the central nervous system, showed the heaviest negative weight in our 38-metabolite score. Regarding dietary factors, we previously, reported that plasma GABA levels were inversely associated with dietary glycemic load intake in a subset of PREDIMED participants [41]. Blood GABA levels were also positively associated with physical activity and bone density in a cross-sectional study in women [42]. In our data, GABA was positively associated ($p = 0.001$) with quintiles of physical activity and this association remained significant after adjusting for age and sex (adjusted beta = 0.048, 95 % CI: 0.014 to 0.081 SD, per additional quintile). Compared to glutamate (a precursor of GABA), neuroprotective effects have been reported for GABA in animal models, particularly after brain or spinal injury [43]. In humans, cortical GABA levels were negatively correlated with depression scores [44].

L-homoarginine (LHA) is an endogenous non-proteinogenic amino acid involved in preserving the endothelial function and consequently may affect the pathogenesis of cardiac dysfunction [45]. Consistent with a previous report from the WHI, we found reduced mortality with higher plasma levels of LHA [6]. LHA is believed to exert protection on vascular function and against oxidative damage, partly because it is an alternative substrate for nitric oxide (NO) synthase and contributes to the generation of the potent vasodilator NO. LHA was related to lower risk of events in patients with cerebrovascular disease and to lower risk of heart failure in the PREDIMED study [46]. Importantly, several studies reported inverse associations of LHA levels with all-cause mortality, particularly in populations with cardiovascular disease or other chronic diseases [47].

Sirtuins, associated with protection against atherosclerosis and cellular senescence leading to longer survival [48], convert NAD(+) into nicotinamide (NAM). We found that higher plasma levels of a metabolite of NAM, 1-Methylnicotinamide, were associated with reduced mortality. To our knowledge, this is a novel finding which was not previously reported. However, it is consistent with the attributed immunomodulatory properties for 1-Methylnicotinamide, including a reduction in reactive oxygen species (particularly superoxide radical anion and hydroxyl radical), an attenuation of the inflammasome, favorable interactions with lipid targets and increases in endothelial prostacyclin (PGI2) and NO, inducing vasorelaxation [49].

CEs and PC-plasmalogens were reported to be consistently associated with a healthy lifestyle score in 4 US cohorts [50]. Additionally, previous reports from the PREDIMED study indicated an inverse association of these metabolites with CVD [51] and T2D [52]. In agreement with these previous findings, we identified some molecules of these families (18 CE:0, 18 CE:2, 20 CE:4, 22 CE:6; PC-plasmalogenB 36:5, PC-plasmalogen 38:6 and PC-plasmalogen 38:7) as negatively weighted in our score, representing predictors of a lower mortality rates.

The plasma metabolites associated with higher mortality were choline (the most heavily positively weighted metabolite in the score), short-chain or long-chain acylcarnitines (CAR 4:0;OH and CAR 18:1;OH, also heavily weighted in the score), other carnitines, 4-acetamidobutanoate, pseudouridine, 7-methylguanine, N6-acetyllysine, DMGV, phenylacetylglutamine and creatinine.

Substantial derangements in acylcarnitines are considered to be associated with disorders of mitochondrial fatty acid oxidation and organic acidemias [53]. Higher plasma levels of acylcarnitines have been consistently associated with higher T2D risk [54], heart failure [55] and mortality in patients with CVD. Concurring with these findings, we identified several acylcarnitines (CAR 4:0;OH, CAR 18:1;OH) and carnitines (CAR 12:1, CAR 14:1, CAR 18:1, CAR 12:0, CAR 10:0) as predictors of a shorter lifespan [56] and worse cardiometabolic health [57].

In consistency with previous findings, we also identified serine associated with mortality, whereas 4-acetamidobutanoate, 1-methylguanosine and pseudouridine were associated with higher risk of mortality [8]. These 3 metabolites (4-acetamidobutanoate, 1-methylguanosine and pseudouridine) were previously reported to be associated with harmful dietary exposures, specifically, sugar-sweetened beverages [23]. In our data, they showed significant age and sex-adjusted partial correlations with BMI and even stronger partial correlations with waist-to-height ratio.

Plasma DMGV was previously associated with liver fat in the Framingham Heart Study Gen III cohort [58]. It was also associated with higher all-cause mortality in individuals with type 2 diabetes in the Mälmo cohort [59]. Furthermore, this metabolite has been reported to be useful to capture an overall unhealthy lifestyle [60]. In our cohort, DMGV was a powerful predictor of earlier death and was directly associated with both BMI (Pearson $r = 0.22$) and waist-to-height ratio (Pearson $r = 0.20$), inversely associated with levels of physical activity (p value $= 0.002$) and directly associated with current smoking (p value $= 0.008$ after adjusting for BMI, age and sex).

All the former considerations contribute to provide a high biological plausibility for this linear combination of plasma metabolites reflecting pathophysiological mechanisms likely related to cardiometabolic processes that may be leading to death.

The main limitation of our study relies on the fact that all PREDIMED participants were subjects at high CVD risk, with a sizable proportion of them having type 2 diabetes (or prediabetes) and hypertension at baseline. This fact imposes a limit in the generalizability of our results. Furthermore, the sample selection was based on case-cohort designs with oversampling of cases (participants developing T2D or CVD during the follow-up period), and we acknowledge that the estimates in the US replication cohorts were significantly lower than in the PREDIMED development cohort. The particular population characteristics of the derivation sample may have contributed to find stronger associations of the metabolite score with mortality in PREDIMED than in other lower-risk cohorts. Another potential limitation is that the identification of metabolites was done in a subset of the baseline PREDIMED data, and a subset of these data was also used to test this prediction. This design may seem to entail the risk of overfitting. Nevertheless, we avoided this risk by using penalized elastic net regression on only 90 % of the sample for the identification of metabolites (and derivation of their coefficients), and thereafter we applied this metabolite selection (weighted by their coefficients) only to the 10 % of the sample that was not used in the derivation set (i.e, the left-out fold). We repeated this procedure ten times to ensure that no subject was used simultaneously for derivation and testing. Additionally, we conducted internal replications using metabolites measured after one year in PREDIMED for the 38 selected metabolites from an elastic net regression. However, for external validation, 27 out of 38 metabolites were available in the NHS-I, NHS-II and HPFS, and 32 out of 38 were available in the WHI, which may have influenced the results. Particularly, in the Cox model of PREDIMED

using individual metabolites (adjusted for each other) to predict all-cause mortality, the following metabolites were significant independent predictors, but they were missing in the NHS/HPFS: Choline ($p = 0.002$); GABA ($p = 0.003$); Guanine ($p = 0.005$); DMGV ($p = 0.016$); trimethilbenzene ($p = 0.049$). Nevertheless, when we reconstructed the score using only the set of metabolites available in the US cohorts, the results were also replicated. The lower values of the HRs estimated in the US cohorts can also be attributed to the unavailability of some metabolites. Lastly, we only assessed identified plasma metabolites, with the potential to restrict our findings only to available metabolites. Yet unidentified metabolite peaks measured using untargeted metabolomic profiling is warranted in further studies to complement the present assessment.

The main strengths of the present analyses include the use of the PREDIMED as prospective cohort with subjects at high cardiovascular risk, presenting a novel alternative for predictors selection. Also, the large sample size, the long-term follow-up, the high predictive capability of the score beyond conventional risk factors, and replications of our findings in independent cohorts on both sides of the Atlantic, thus confirming the external validity of our results. The utilization of repeated metabolomic measurements in PREDIMED at baseline and 1-year follow-up allows us to assess the stability of metabolomic signatures over time. In addition, we utilized robust methods. Elastic net regression modelling used in our identification of the predictive score is known to be best suited for feature selection in high-dimensional data with high correlations among predictors [29].

In conclusion, we identified plasma multi-metabolite profiles able to predict overall mortality, and especially, earlier mortality. Scores based on identified metabolites could be further investigated to provide mechanistic explanations for their causal role on overall survival and for their clinical use in the context of precision medicine, particularly in patients with T2D or at a high risk of CVD.

CRediT authorship contribution statement

Gonzalo Fernández-Duval: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Cristina Razquin:** Writing – review & editing, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Fenglei Wang:** Writing – original draft, Validation, Methodology. **Huan Yun:** Writing – original draft, Validation. **Jie Hu:** Writing – original draft, Validation. **Marta Guasch-Ferré:** Writing – review & editing, Validation. **Kathryn Rexrode:** Writing – review & editing, Validation, Project administration, Investigation. **Raji Balasubramanian:** Writing – review & editing, Validation. **Jesús García-Gavilán:** Writing – review & editing. **Miguel Ruiz-Canela:** Writing – review & editing, Project administration. **Clary B. Clish:** Writing – review & editing, Data curation. **Dolores Corella:** Writing – review & editing, Project administration. **Enrique Gómez-Gracia:** Writing – review & editing, Project administration. **Miquel Fiol:** Writing – review & editing, Project administration. **Ramón Estruch:** Writing – review & editing, Project administration. **José Lapetra:** Writing – review & editing, Project administration. **Montse Fitó:** Writing – review & editing, Project administration. **Luis Serra-Majem:** Writing – review & editing, Project administration. **Emilio Ros:** Writing – review & editing, Project administration. **Liming Liang:** Writing – review & editing, Methodology. **Courtney Dennis:** Writing – review & editing. **Eva M. Asensio:** Writing – review & editing. **Olga Castañer:** Writing – review & editing, Project administration. **Francisco J. Planes:** Writing – review & editing, Methodology. **Jordi Salas-Salvadó:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization, Project administration. **Frank B. Hu:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition, Conceptualization. **Estefanía Toledo:** Writing – review & editing, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Miguel A. Martínez-González:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Funding sources

The PREDIMED trial was funded by the Spanish Ministry of Health (Instituto de Salud Carlos III [ISCIII], through the PREDIMED Network grant RD 06/0045, 2006–2013, coordinated by MAM-G; and a previous network grant RTIC-G03/140, 2003–2005, coordinated by 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 by CIBER Fisiopatología de la Obesidad y Nutrición (CIBERobn). The NHS, NHSII, HPFS, and PREDIMED are supported by grants from the National Institutes of Health through contracts UM1 CA186107, U01 CA176726, U01 CA167552, U01 HL145386, R01 CA49449, R01 HL034594, R01 HL088521, R01 CA67262, R01 HL35464, R01 HL60712, R01 CA50385, P01 CA87969, R01 AR049880, R01 DK112940, R01 DK119268, R01 DK120870, P30 DK046200, R01 HL118264, and R01 DK127601. The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts 75N92021D00001, 75N92021D00002, 75N92021 D00003, 75N92021D00004, 75N92021D00005, and metabolomic measures through HHSN268201300008C. Finally, the authors would like to acknowledge the contribution to NHS-I, NHS-II, and HPFS from central cancer registries supported through the Centers for Disease Control and Prevention's National Program of Cancer Registries (NPCR) and/or the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program. Central registries may also be supported by state agencies, universities, and cancer centers. Participating central cancer registries include the following: Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Indiana, Iowa, Kentucky, Louisiana, Massachusetts, Maine, Maryland, Michigan, Mississippi, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Puerto Rico, Rhode Island, Seattle SEER Registry, South Carolina, Tennessee, Texas, Utah, Virginia, West Virginia, Wyoming.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Jordi Salas-Salvado reports a relationship with International Nut and Dried Fruit Foundation that includes: funding grants and travel reimbursement. Jordi Salas-Salvado reports a relationship with Mundipharma that includes: speaking and lecture fees. Jordi Salas-Salvado reports a relationship with International Advisory Board of the Project Effect of cashew nut supplementation on glycemic status and lipid profile in type 2 diabetes subjects that includes: board membership. Jordi Salas-Salvado reports a relationship with Institute Danone Spain Advisory Board that includes: board membership. Jordi Salas-Salvado reports a relationship with Scientific Committee of Danone Institute International that includes: board membership. Jordi Salas-Salvado reports a relationship with International Nut and Dried Fruit Foundation World Forum for Nutrition Research and Dissemination that includes: board membership. Jordi Salas-Salvadó is partially supported by ICREA under the ICREA Academia programme. Ramon Estruch reports a relationship with Sociedad Española de Nutrición that includes: non-financial support. Ramon Estruch reports a relationship with Fundación Bosch y Gimpera that includes: non-financial support. Ramon Estruch reports a relationship with Brewers of Europe that includes: speaking and lecture fees. Ramon Estruch reports a relationship with Fundación Cerveza y Salud that includes: speaking and lecture fees.

Ramon Estruch reports a relationship with Interprofesional del Aceite de Oliva that includes: speaking and lecture fees. Ramon Estruch reports a relationship with Instituto Cervantes that includes: speaking and lecture fees. Ramon Estruch reports a relationship with Pernaud Ricard that includes: speaking and lecture fees. Ramon Estruch reports a relationship with Fundación Dieta Mediterránea that includes: speaking and lecture fees. Ramon Estruch reports a relationship with Wine and Culinary International Forum that includes: speaking and lecture fees. Ramon Estruch reports a relationship with Fundación Patrimonio Comunal Olivarero that includes: non-financial support. Emilio Ros reports a relationship with California Walnut Commission that includes: speaking and lecture fees. Emilio Ros reports a relationship with Alexion that includes: consulting or advisory. Emilio Ros reports a relationship with International Nut and Dried Fruit Council that includes: travel reimbursement. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors appreciate and thank all the PREDIMED, NHS-I, NHS-II, HPFS and WHI participants and staff for the collaboration.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.metabol.2025.156195.

Code availability

The main code of the analyses presented in this study are available on GitHub: https://github.com/gfernandezduval/metabolomicscore_mortality/.

References

[1] Balasubramanian R, Shutta KH, Guasch-Ferre M, Huang T, Jha SC, Zhu Y, et al. Metabolomic profiles of chronic distress are associated with cardiovascular disease risk and inflammation-related risk factors. Brain Behav Immun 2023;114:262–74. https://doi.org/10.1016/J.BBI.2023.08.010.

[2] Martínez-González M, Ruiz-Canela M, Hruby A, Liang L, Trichopoulou A, Hu FB. Intervention trials with the Mediterranean diet in cardiovascular prevention: understanding potential mechanisms through Metabolomic profiling. J Nutr 2016;146:913S–9S. https://doi.org/10.3945/JN.115.219147.

[3] Zhang H, Zhou XD, Shapiro MD, Lip GYH, Tilg H, Valenti L, et al. Global burden of metabolic diseases, 1990–2021. Metabolism 2024;160:155999. https://doi.org/10.1016/J.METABOL.2024.155999.

[4] Buergel T, Steinfeldt J, Ruyoga G, Pietzner M, Bizzarri D, Vojinovic D, et al. Metabolomic profiles predict individual multidisease outcomes. Nat Med 2022;28:2309–20. https://doi.org/10.1038/s41591-022-01980-3.

[5] López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. Hallmarks of aging: an expanding universe. Cell 2023;186:243–78. https://doi.org/10.1016/J.CELL.2022.11.001.

[6] Balasubramanian R, Paynter NP, Giulianini F, Manson JE, Zhao Y, Chen JC, et al. Metabolomic profiles associated with all-cause mortality in the Women's Health Initiative. Int J Epidemiol 2020;49:289–300. https://doi.org/10.1093/IJE/DYZ211.

[7] Yan Y, Smith E, Melander O, Ottosson F. The association between plasma metabolites and future risk of all-cause mortality. J Intern Med 2022;292:804–15. https://doi.org/10.1111/JOIM.13540.

[8] Wang F, Tessier A-J, Liang L, Wittenbecher C, Haslam DE, Fernández-Duval G, et al. Plasma metabolomic profiles associated with mortality and longevity in a prospective analysis of 13,512 individuals. Nat Commun 2023;14:5744. https://doi.org/10.1038/s41467-023-41515-z.

[9] Ndrepepa G, Braun S, King L, Hadamitzky M, Haase HU, Birkmeier KA, et al. Association of uric acid with mortality in patients with stable coronary artery disease. Metabolism 2012;61:1780–6. https://doi.org/10.1016/J.METABOL.2012.05.014.

[10] 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;41:2645–56. https://doi.org/10.1093/eurheartj/ehaa209.

[11] Andreo-López MC, Contreras-Bolívar V, Muñoz-Torres M, García-Fontana B, García-Fontana C. Influence of the Mediterranean diet on healthy aging. Int J Mol Sci 2023;24:4491. https://doi.org/10.3390/IJMS24054491.

[12] Estruch R, Ros E, Salas-Salvadó J, Covas M-I, 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:e34. https://doi.org/10.1056/NEJMOA1800389/SUPPL_FILE/NEJMOA1800389_DISCLOSURES.PDF.

[13] Martínez-González MA, Salas-Salvadó J, Estruch R, Corella D, Fitó M, Ros E. Benefits of the Mediterranean diet: insights from the PREDIMED study. Prog Cardiovasc Dis 2015;58:50–60. https://doi.org/10.1016/J.PCAD.2015.04.003.

[14] Bao Y, Bertoia ML, Lenart EB, Stampfer MJ, Willett WC, Speizer FE, et al. Origin, methods, and evolution of the three nurses' health studies. Am J Public Health 2016;106:1573–81. https://doi.org/10.2105/AJPH.2016.303338.

[15] Hankinson SE, Willett WC, Manson JAE, Colditz GA, Hunter DJ, Spiegelman D, et al. Plasma sex steroid hormone levels and risk of breast Cancer in postmenopausal women. JNCI J Natl Cancer Inst 1998;90:1292–9. https://doi.org/10.1093/JNCI/90.17.1292.

[16] Platz EA, Clinton SK, Giovannucci E. Association between plasma cholesterol and prostate cancer in the PSA era. Int J Cancer 2008;123:1693–8. https://doi.org/10.1002/IJC.23715.

[17] Anderson G, Cummings S, Freedman LS, Furberg C, Henderson M, Johnson SR, et al. Design of the Women's health initiative clinical trial and observational study. Control Clin Trials 1998;19:61–109. https://doi.org/10.1016/S0197-2456(97)00078-0.

[18] Anderson GL, Limacher M. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. JAMA 2004;291:1701–12. doi:https://doi.org/10.1001/JAMA.291.14.1701.

[19] Martínez-González MÁ, Investigators for the PS, Corella D, Investigators for the PS, Salas-Salvadó J, Investigators for the PS, et al. Cohort profile: design and methods of the PREDIMED study. Int J Epidemiol 2012;41:377–85. doi:https://doi.org/10.1093/IJE/DYQ250.

[20] 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;103:1408–16. https://doi.org/10.3945/AJCN.116.130492.

[21] Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis: chemical analysis working group (CAWG) metabolomics standards initiative (MSI). Metabolomics 2007;3:211–21. https://doi.org/10.1007/S11306-007-0082-2/METRICS.

[22] 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;137:841–53. https://doi.org/10.1161/CIRCULATIONAHA.117.029468/-/DC1.

[23] Wang F, Baden MY, Guasch-Ferré M, Wittenbecher C, Li J, Li Y, et al. Plasma metabolite profiles related to plant-based diets and the risk of type 2 diabetes. Diabetologia 2022;65:1119–32. https://doi.org/10.1007/S00125-022-05692-8.

[24] Hu J, Yao J, Deng S, Balasubramanian R, Jiménez MC, Li J, et al. Differences in Metabolomic profiles between black and white women and risk of coronary heart disease: an observational study of women from four US cohorts. Circ Res 2022;131:601–15. https://doi.org/10.1161/CIRCRESAHA.121.320134.

[25] 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;8:1–10. https://doi.org/10.1038/s41598-017-19120-0.

[26] Schröder H, Fitó M, Estruch R, Martínez-González MA, Corella D, Salas-Salvadó J, et al. A short screener is valid for assessing Mediterranean diet adherence among older Spanish men and women. J Nutr 2011;141:1140–5. https://doi.org/10.3945/JN.110.135566.

[27] Al-Shaar L, Yuan C, Rosner B, Dean SB, Ivey KL, Clowry CM, et al. Reproducibility and validity of a Semiquantitative food frequency questionnaire in men assessed by multiple methods. Am J Epidemiol 2021;190:1122–32. https://doi.org/10.1093/AJE/KWAA280.

[28] Patterson RE, Kristal AR, Tinker LF, Carter RA, Bolton MP, Agurs-Collins T. Measurement characteristics of the Women's Health Initiative food frequency questionnaire. Ann Epidemiol 1999;9:178–87. https://doi.org/10.1016/S1047-2797(98)00055-6.

[29] Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B Stat Methodol 2005;67:301–20. https://doi.org/10.1111/J.1467-9868.2005.00503.X.

[30] Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. J Stat Softw 2011;39:1–13. https://doi.org/10.18637/JSS.V039.I05.

[31] Kernbach JM, Staartjes VE. Foundations of machine learning-based clinical prediction modeling: part II—generalization and overfitting. Acta Neurochir Suppl (Wien) 2022;134:15–21. https://doi.org/10.1007/978-3-030-85292-4_3/FIGURES/2.

[32] Discacciati A, Bellavia A, Orsini N, Greenland S. On the interpretation of risk and rate advancement periods. Int J Epidemiol 2016;45:278–84. https://doi.org/10.1093/IJE/DYV320.

[33] Brenner H, Gefeller O, Greenland S. Risk and rate advancement periods as measures of exposure impact on the occurrence of chronic diseases. Epidemiology 1993;4:229–36. https://doi.org/10.1097/00001648-199305000-00006.

[34] Brenner H, Gefeller O, Greenland S. Risk and rate advancement periods as measures of exposure impact on the occurrence of chronic diseases. Epidemiology 1993;4:229–36. https://doi.org/10.1097/00001648-199305000-00006.

[35] Manzo OL, Nour J, Sasset L, Marino A, Rubinelli L, Palikhe S, et al. Rewiring endothelial sphingolipid metabolism to favor S1P over ceramide protects from coronary atherosclerosis. Circ Res 2024;134:990–1005. https://doi.org/10.1161/CIRCRESAHA.123.323826.

[36] Nikolova-Karakashian M, Karakashian A, Rutkute K. Role of neutral sphingomyelinases in aging and inflammation. Subcell Biochem 2008;49:469–86. https://doi.org/10.1007/978-1-4020-8831-5_18.
[37] Fretts AM, Jensen PN, Sitlani CM, Hoofnagle A, Lidgard B, Umans JG, et al. Circulating sphingolipids and all-cause mortality: the strong heart family study. J Am Heart Assoc 2024;13:e032536. https://doi.org/10.1161/JAHA.123.032536/SUPPL_FILE/JAH39776-SUP-0001-SUPINFO.PDF.
[38] 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;45:1013–24. https://doi.org/10.2337/DC21-1705.
[39] Wang DD, Toledo E, Hruby A, Rosner BA, Willett WC, Sun Q, et al. Plasma ceramides, Mediterranean diet, and incident cardiovascular disease in the PREDIMED trial (Prevención con Dieta Mediterránea). Circulation 2017;135:2028–40. https://doi.org/10.1161/CIRCULATIONAHA.116.024261.
[40] Li D, Misialek JR, Huang F, Windham GB, Yu F, Alonso A. Independent Association of Plasma Hydroxysphingomyelins with Physical Function in the atherosclerosis risk in communities (ARIC) study. J Gerontol A Biol Sci Med Sci 2018;73:1103–10. https://doi.org/10.1093/GERONA/GLX201.
[41] Sasset L, Di Lorenzo A. Sphingolipid metabolism and signaling in endothelial cell functions. Adv Exp Med Biol 2022;1372:87–117. https://doi.org/10.1007/978-981-19-0394-6_8.
[42] Wang Z, Bian L, Mo C, Shen H, Zhao LJ, Su KJ, et al. Quantification of aminobutyric acids and their clinical applications as biomarkers for osteoporosis. Commun Biol 2020;3:39. https://doi.org/10.1038/S42003-020-0766-Y.
[43] Llorente IL, Perez-Rodriguez D, Martínez-Villayandre B, Dos-Anjos S, Darlison MG, Poole AV, et al. GABA(a) receptor chloride channels are involved in the neuroprotective role of GABA following oxygen and glucose deprivation in the rat cerebral cortex but not in the hippocampus. Brain Res 2013;1533:141–51. https://doi.org/10.1016/J.BRAINRES.2013.08.024.
[44] Wang J, Liu G, Xu K, Ai K, Huang W, Zhang J. The role of neurotransmitters in mediating the relationship between brain alterations and depressive symptoms in patients with inflammatory bowel disease. Hum Brain Mapp 2023;44:5357–71. https://doi.org/10.1002/HBM.26439.
[45] Koch V, Gruenewald LD, Gruber-Rouh T, Eichler K, Leistner DM, Mahmoudi S, et al. Homoarginine in the cardiovascular system: pathophysiology and recent developments. Fundam Clin Pharmacol 2023;37:519–29. https://doi.org/10.1111/FCP.12858.
[46] Goni L, Razquin C, Toledo E, Guasch-Ferré M, Clish CB, Babio N, et al. Arginine catabolism metabolites and atrial fibrillation or heart failure risk: 2 case-control studies within the Prevención con Dieta Mediterránea (PREDIMED) trial. Am J Clin Nutr 2022;116:653–62. https://doi.org/10.1093/AJCN/NQAC139.
[47] Mokhaneli MC, Botha-Le Roux S, Fourie CMT, Böger R, Schwedhelm E, Mels CMC. L-homoarginine is associated with decreased cardiovascular- and all-cause mortality. Eur J Clin Investig 2021;51:e13472. https://doi.org/10.1111/ECI.13472.
[48] Sazdova I, Hadzi-Petrushev N, Keremidarska-Markova M, Stojchevski R, Sopi R, Shileiko S, et al. SIRT-associated attenuation of cellular senescence in Vascular Wall. Mech Ageing Dev 2024;220:111943. https://doi.org/10.1016/J.MAD.2024.111943.
[49] Sidor K, Jeznach A, Hoser G, Skirecki T. 1-Methylnicotinamide (1-MNA) inhibits the activation of the NLRP3 inflammasome in human macrophages. Int Immunopharmacol 2023;121:110445. https://doi.org/10.1016/J.INTIMP.2023.110445.
[50] Tessier AJ, Wang F, Liang L, Wittenbecher C, Haslam DE, Eliassen AH, et al. Plasma metabolites of a healthy lifestyle in relation to mortality and longevity: four prospective US cohort studies. Med 2024;5:224–238.e5. https://doi.org/10.1016/J.MEDJ.2024.01.010.
[51] Razquin C, Liang L, Toledo E, Clish CB, Ruiz-Canela M, Zheng Y, et al. Plasma lipidome patterns associated with cardiovascular risk in the PREDIMED trial: a case-cohort study. Int J Cardiol 2018;253:126–32. https://doi.org/10.1016/J.IJCARD.2017.10.026.
[52] Razquin C, Toledo E, Clish CB, Ruiz-Canela M, Dennis C, Corella D, et al. Plasma Lipidomic profiling and risk of type 2 diabetes in the PREDIMED trial. Diabetes Care 2018;41:2617–24. https://doi.org/10.2337/DC18-0840.
[53] Driscoll DA, Gross SJ. Acylcarnitine profile analysis. Genet Med 2008;10:156. https://doi.org/10.1097/GIM.0B013E3181614289.
[54] 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;104:1508–19. https://doi.org/10.1210/JC.2018-01000.
[55] 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. Rev Esp Cardiol (Engl Ed) 2022;75:649–58. https://doi.org/10.1016/J.REC.2021.10.005.
[56] Wang T, Beyene HB, Yi C, Cinel M, Mellett NA, Olshansky G, et al. A lipidomic based metabolic age score captures cardiometabolic risk independent of chronological age. EBioMedicine 2024;105:105199. https://doi.org/10.1016/J.EBIOM.2024.105199.
[57] Batch BC, Shah SH, Newgard CB, Turer CB, Haynes C, Bain JR, et al. Branched chain amino acids are novel biomarkers for discrimination of metabolic wellness. Metabolism 2013;62:961–9. https://doi.org/10.1016/J.METABOL.2013.01.007.
[58] 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 Invest 2017;127:4394–402. https://doi.org/10.1172/JCI95995.
[59] Ottosson F, Smith E, Fernandez C, Melander O. Plasma metabolites associate with all-cause mortality in individuals with type 2 diabetes. Metabolites 2020;10:1–11. https://doi.org/10.3390/METABO10080315.
[60] Rios S, García-Gavilán JF, Babio N, Paz-Graniel I, Ruiz-Canela M, Liang L, et al. Plasma metabolite profiles associated with the World Cancer Research Fund/American Institute for Cancer Research lifestyle score and future risk of cardiovascular disease and type 2 diabetes. Cardiovasc Diabetol 2023;22:252. https://doi.org/10.1186/S12933-023-01912-6.