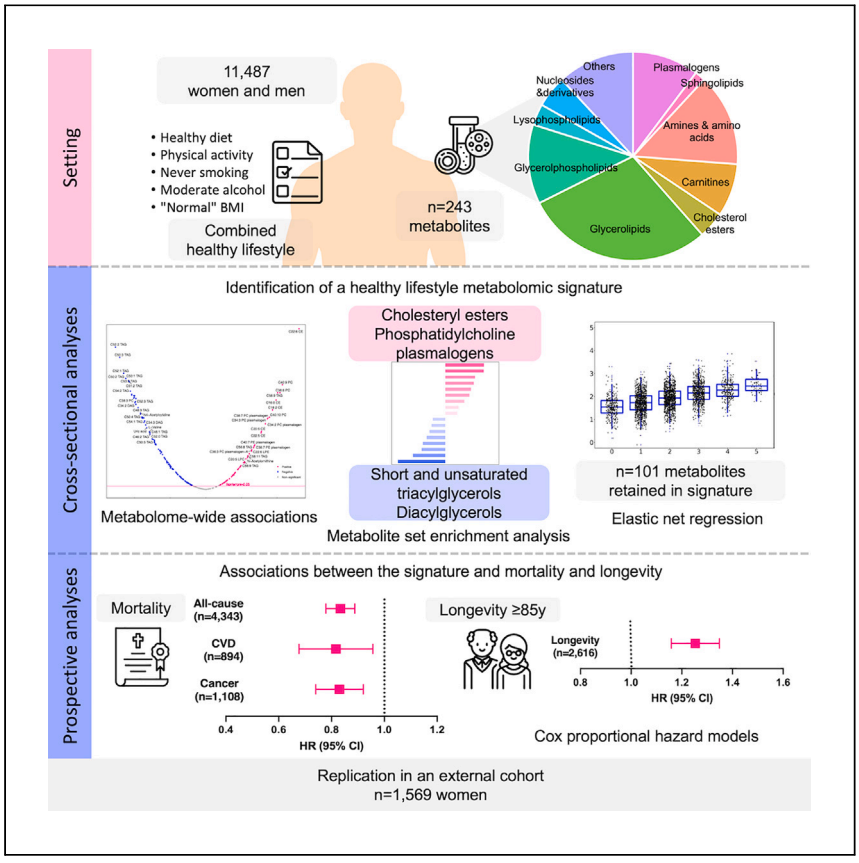
Plasma metabolites of a healthy lifestyle in relation to mortality and longevity: Four prospective US cohort studies

Med

CellPress
OPEN ACCESS

Clinical and Translational Article

Plasma metabolites of a healthy lifestyle in relation to mortality and longevity: Four prospective US cohort studies



In four large US cohorts including more than 13,000 adults, a unique metabolomic signature of a healthy lifestyle was identified and replicated. The signature is highly reflective of lipid metabolism pathways and is associated with lower mortality risk and greater likelihood of longevity over 28 years of follow-up.

Anne-Julie Tessier, Fenglei Wang, Liming Liang, ..., Jorge E. Chavarro, Frank B. Hu, Marta Guasch-Ferré

ajtessier@hsph.harvard.edu (A.-J.T.)
mguasch@hsph.harvard.edu (M.G.-F.)

Highlights

Greater adherence to a healthy lifestyle may lead to changes in the metabolome

A healthy lifestyle metabolomic signature is reflective of lipid metabolism pathways

A higher signature score is associated with lower mortality risk and greater longevity

The signature mediates the association of a healthy lifestyle with mortality and longevity

Translation to Population Health

Tessier et al., Med 5, 224–238
March 8, 2024
https://doi.org/10.1016/j.medj.2024.01.010

CellPress
OPEN ACCESS

Med

Clinical and Translational Article

Plasma metabolites of a healthy lifestyle in relation to mortality and longevity: Four prospective US cohort studies

Anne-Julie Tessier,[1,*] Fenglei Wang,[1] Liming Liang,[2,3] Clemens Wittenbecher,[4] Danielle E. Haslam,[1,5] A. Heather Eliassen,[1,2,5] Deirdre K. Tobias,[1,6] Jun Li,[1,6] Oana A. Zeleznik,[5] Alberto Ascherio,[1,5] Qi Sun,[1,5] Meir J. Stampfer,[1,2,5] Francine Grodstein,[2,7] Kathryn M. Rexrode,[8] JoAnn E. Manson,[2,5,6] Raji Balasubramanian,[9] Clary B. Clish,[10] Miguel A. Martínez-González,[1,11] Jorge E. Chavarro,[1,2,5] Frank B. Hu,[1,2,5] and Marta Guasch-Ferré[1,12,13,*]

SUMMARY

Background: A healthy lifestyle is associated with a lower premature mortality risk and with longer life expectancy. However, the metabolic pathways of a healthy lifestyle and how they relate to mortality and longevity are unclear. We aimed to identify and replicate a healthy lifestyle metabolomic signature and examine how it is related to total and cause-specific mortality risk and longevity.

Methods: In four large cohorts with 13,056 individuals and 28-year follow-up, we assessed five healthy lifestyle factors, used liquid chromatography mass spectrometry to profile plasma metabolites, and ascertained deaths with death certificates. The unique healthy lifestyle metabolomic signature was identified using an elastic regression. Multivariable Cox regressions were used to assess associations of the signature with mortality and longevity.

Findings:: The identified healthy lifestyle metabolomic signature was reflective of lipid metabolism pathways. Shorter and more saturated triacylglycerol and diacylglycerol metabolite sets were inversely associated with the healthy lifestyle score, whereas cholesteryl ester and phosphatidylcholine plasmalogen sets were positively associated. Participants with a higher healthy lifestyle metabolomic signature had a 17% lower risk of all-cause mortality, 19% for cardiovascular disease mortality, and 17% for cancer mortality and were 25% more likely to reach longevity. The healthy lifestyle metabolomic signature explained 38% of the association between the self-reported healthy lifestyle score and total mortality risk and 49% of the association with longevity.

Conclusions: This study identifies a metabolomic signature that measures adherence to a healthy lifestyle and shows prediction of total and cause-specific mortality and longevity.

Funding: This work was funded by the NIH, CIHR, AHA, Novo Nordisk Foundation, and SciLifeLab.

CONTEXT AND SIGNIFICANCE

Eating well, staying active, not smoking, and maintaining a healthy body weight are well known to protect against diseases and early death. However, it is not clear whether adhering to these healthy habits affects specific molecules in the body. In over 13,000 adults, Tessier et al. demonstrated that adherence to a healthy lifestyle is associated with the level of molecules in the blood, particularly those in the body's processing of fats. The researchers identified a unique combination of molecules that explains part of the association between greater adherence to a healthy lifestyle and lower risk of death and higher likelihood of living longer. The study provides new biological understanding about how living healthily is linked to lower risk of dying early.

INTRODUCTION

A combination of healthy lifestyle factors including a healthy diet, body mass index (BMI) within 18.5–24.9 kg/m^2, moderate alcohol intake, higher levels of physical activity, and never smoking was associated with 55%–71% lower risk of all-cause mortality in prospective cohort studies.[1,2] In US populations, adhering to all low-risk

lifestyle factors was shown to increase life expectancy by up to 14 years compared to adhering to none.[3] Nonetheless, the precise mechanisms linking lifestyle behaviors with mortality and longevity are not well understood.

High-throughput metabolite profiling provides a comprehensive picture of individual metabolic status and reflects the intricate interplay between lifestyle factors and genetic susceptibility.[4] Advances in metabolomic techniques hold promise in providing insights into underlying biological mechanisms, identifying high-risk individuals, and developing targeted preventive lifestyle interventions to reduce premature mortality.

Previous studies have suggested that metabolomic signatures may be associated with body weight,[5] diet,[6] physical activity,[7] alcohol consumption,[8,9] and smoking.[10,11] Yet, few studies have examined metabolites associated with a composite healthy lifestyle score.[12–15] One systematic review recently examined the literature on the adherence to a healthy lifestyle, as defined by two or more combined healthy lifestyle behaviors, and metabolomics in humans (9 studies, n = 34–3,234 participants, observational and trial designs).[16] Although mixed, findings from these studies generally supported positive associations of polyunsaturated fatty acids (PUFAs), phosphatidylcholines (PCs), and some amino acids, mainly glutamate, with a healthy lifestyle and inverse associations of triacylglycerols (TAG), sphingomyelins (SMs), and carnitines, among other metabolites.[16] However, most studies only examined diet and physical activity factors, with small sample sizes and limited sets of metabolites profiled. Thus, a comprehensive understanding of the metabolic pathways underlying healthy lifestyle behaviors remains to be discovered. By studying several modifiable lifestyle factors simultaneously, a better understanding of the common biological mechanisms as well as the key differences may be acquired.

In the present study, by leveraging lifestyle, metabolomic, and clinical data from >13,000 participants in four US adult prospective cohorts (three discovery and one external replication), we aimed to identify a metabolomic signature of a combined healthy lifestyle score in mid-life and examine its prospective association with all-cause and cause-specific mortality and longevity over up to 28 years of follow-up.

RESULTS

Participants' characteristics

A total of 11,487 participants from the Nurses' Health Study (NHS), NHSII, and Health Professionals Follow-up Study (HPFS) were included in our primary analyses (Figure 1). Participants were predominantly middle-aged (mean age 54.3 ± 9.0 years) women (85.8%) of White ethnicity (96.7%) and reported adhering to 1.9 ± 1.1 healthy lifestyle factors (Table 1). A total of 1,569 participants from the Women's Health Initiative (WHI) study were included in the replication analyses. The WHI participants were older, had a higher prevalence of diabetes, hypertension, and hypercholesterolemia, and reported lower adherence to healthy lifestyle factors (mean 1.5 ± 1.1) (Table 1). In the NHS/NHSII/HPFS, participants who adhered to 5 out of 5 healthy lifestyle factors were most likely to be men; they had a lower BMI, higher alcohol intake, and higher diet quality; they never smoked; and they exercised more compared to those adhering to fewer factors; they also had a lower prevalence of diabetes, hypertension, and hypercholesterolemia (Table S1).

Metabolome-wide associations

After adjusting for covariates, 58 metabolites were positively associated and 129 were inversely associated with the self-reported healthy lifestyle score (Bonferroni-adjusted

[1]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[2]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[3]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[4]Chalmers University of Technology, Gothenburg, Sweden
[5]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
[6]Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
[7]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA
[8]Division of Women's Health, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
[9]Department of Biostatistics and Epidemiology, University of Massachusetts Amherst, Amherst, MA, USA
[10]Broad Institute of MIT and Harvard, Cambridge, MA, USA
[11]Department of Preventive Medicine and Public Health, University of Navarra, Pamplona, Spain
[12]Department of Public Health and Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[13]Lead contact
*Correspondence: ajtessier@hsph.harvard.edu (A.-J.T.), mguasch@hsph.harvard.edu (M.G.-F.)
https://doi.org/10.1016/j.medj.2024.01.010

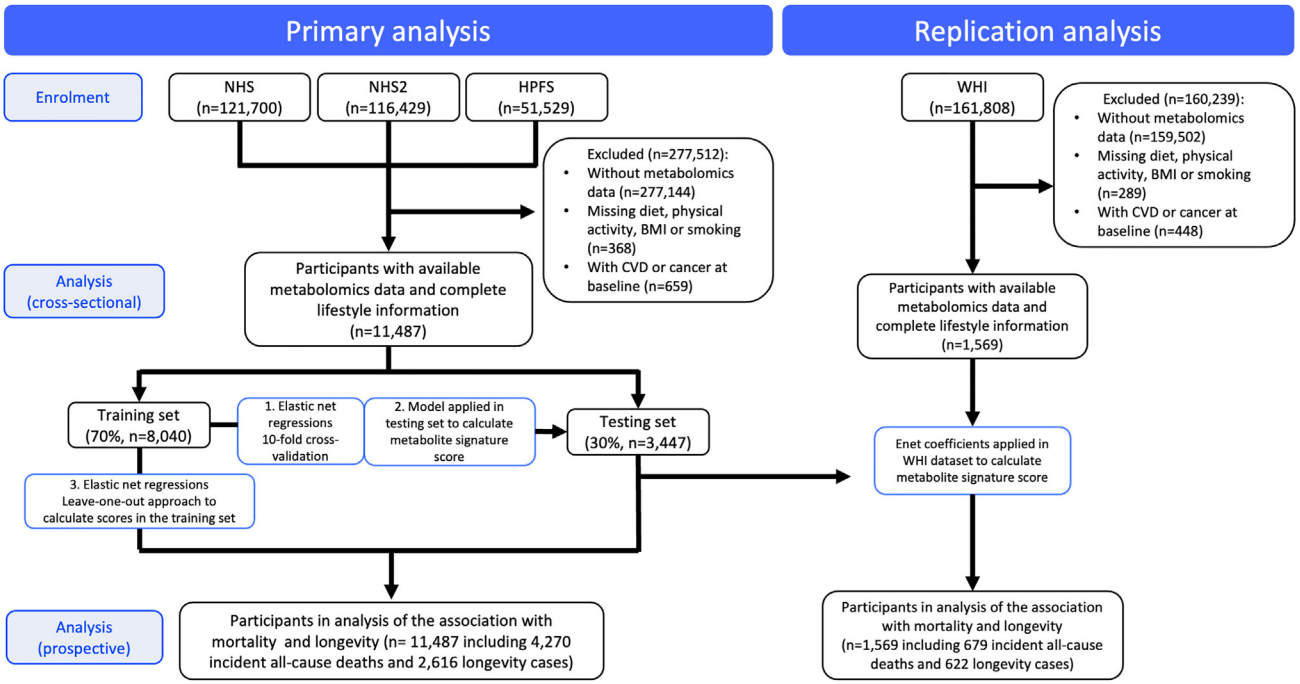


Figure 1. Flow diagram

NHS, Nurses' Health Study I; NHSII, Nurses' Health Study II; HPFS, Health Professionals Follow-up Study; BMI, body mass index; CVD, cardiovascular disease.

$p < 0.05$). The most strongly positively associated metabolites included C22:6 and C18:cholesteryl ester (CE) 2, C58:9 triacylglycerol (TAG), C40:9 and C38:6 PCs, several other CEs, and phospholipid plasmalogens (PCs and phosphatidylethanolamines [PEs]), and the most strongly inversely associated were predominantly TAGs with shorter chain and higher saturation (namely C52:2, C50:3, C50:2, C50:1) and diacylglycerols (DAGs; namely C32:1 and C34:1) (Figure 2A; Table S2). Results remained consistent across age groups (<55 vs. ≥55 years; Pearson r for beta coefficients = 0.95, r^2 = 0.90), sex (women vs. men; Pearson r for beta coefficients = 0.90, r^2 = 0.81), and case-control status (case vs. control; Pearson r for beta coefficients = 0.98, r^2 = 0.96). From metabolite set enrichment analysis (MSEA), four metabolite classes were significantly associated with the self-reported healthy lifestyle score (false discovery rate $p < 0.05$; Figure 2B). CE and PC plasmalogens emerged as the most enriched metabolite groups positively associated with the healthy lifestyle score, whereas TAGs (≤56 carbons and ≤3 double bonds) and DAGs were the most enriched metabolite groups inversely associated.

Results from the individual linear regressions without adjusting for potential biological intermediates were consistent with the main analysis (Table S3). In the sensitivity analysis of the self-reported healthy lifestyle score without the alcohol component, 186 metabolites were significantly associated with the healthy lifestyle score (Bonferroni-adjusted $p < 0.05$). Specifically, 55 were positively associated and 131 were inversely associated. The most strongly associated metabolites consistently included C22:CE 6, C58:9 TAG, and C18:CE 2 (positive associations) and C52:2, C50:1, C50:2, and C52:1 TAGs (inverse associations) (Figure S1A; Table S4). The MSEA showed the same enriched metabolite groups associated with the healthy lifestyle score with and without the alcohol component, i.e., CE and PC plasmalogens (positive associations) and TAGs (≤56 carbons and ≤3 double bonds) and DAGs (inverse associations) (Figure S1B).

Table 1. Baseline characteristics of participants

	NHS	NHSII	HPFS	Overall	WHI
n	6,731	3,121	1,635	11,487	1,569
Age, years	56.8 (6.8)	44.6 (4.5)	62.7 (8.3)	54.3 (9.0)	66.9 (7.1)
Women, n (%)	6,731 (100.0)	3,121 (100.0)	0 (0.0)	9,852 (85.8)	1,569 (100.0)
White, n (%)	6,492 (96.4)	3,036 (97.3)	1,584 (96.9)	11,112 (96.7)	1,244 (79.3)
Fasting at blood collection, n (%)	5,079 (75.5)	2,251 (72.1)	962 (58.8)	8,292 (72.2)	1,569 (100.0)
Multivitamin use, n (%)	4,366 (64.9)	2,390 (76.6)	953 (58.3)	7,709 (67.1)	58 (3.7)
Diabetes, n (%)	222 (3.3)	50 (1.6)	63 (3.9)	335 (2.9)	109 (6.9)
Hypertension, n (%)	1,799 (26.7)	347 (11.1)	452 (27.6)	2,598 (22.6)	384 (24.5)
Hypercholesterolemia, n (%)	1,842 (27.4)	780 (25.0)	591 (36.1)	3,213 (28.0)	159 (10.1)
Hypertensive medication use, n (%)	1,533 (22.8)	244 (7.8)	394 (24.1)	2,171 (18.9)	478 (30.5)
Lipid-lowering medication use, n (%)	163 (2.4)	125 (4.0)	110 (6.7)	398 (3.5)	137 (8.7)
Energy intake, kcal/d	1,777 (469)	1,836 (499)	2,033 (555)	1,830 (498)	1,564 (582)
Healthy lifestyle adherence, 0–5	1.8 (1.1)	2.1 (1.1)	2.2 (1.2)	1.9 (1.1)	1.5 (1.1)
Smoking, n (%)					
Never	3,163 (47.0)	2,084 (66.8)	794 (48.6)	6,041 (52.6)	781 (49.8)
Past	2,792 (41.5)	793 (25.4)	777 (47.5)	4,362 (38.0)	613 (39.1)
Current	776 (11.5)	244 (7.8)	64 (3.9)	1,084 (9.4)	175 (11.2)
BMI, kg/m^2	25.6 (4.6)	25.9 (5.5)	25.8 (3.3)	25.7 (4.7)	28.5 (6.1)
Moderate to vigorous physical activity, h/week	2.2 (3.3)	3.0 (3.3)	4.0 (4.7)	2.7 (3.6)	1.9 (3.1)
Alcohol intake, g/day	5.6 (9.1)	3.9 (6.7)	11.5 (14.2)	6.0 (9.7)	4.7 (10.3)
AHEI, score 0–100	47.9 (9.6)	46.4 (9.8)	48.7 (10.3)	47.6 (9.8)	47.1 (9.7)

Values are means (SD) for continuous variables and counts (%) for categorical variables. Alternative healthy eating index (AHEI) without alcohol. NHS, Nurses' Health Study I; NHSII, Nurses' Health Study II; HPFS, Health Professionals Follow-up Study; SD, standard deviation; BMI, body mass index.

As for individual healthy lifestyle factors, 183 metabolites were associated with a BMI between 18.5 and 24.9 kg/m^2, 152 with a healthier diet, 80 with low-moderate alcohol intake, 35 with moderate to vigorous physical activity level, and 85 with never smoking (all Bonferroni-adjusted $p < 0.05$; Tables S5, S6, S7, S8, and S9). Metabolites that were most positively associated with a "normal" BMI were CE and PC plasmalogens, whereas shorter-chain TAGs ($\leq$56 carbons and $\leq$3 double bonds), uric acid, DAGs, and valine were inversely associated. Long-chain TAGs, PCs, and CEs were the metabolites most positively associated with a healthier diet, whereas myristoleic acid, C7 carnitine, PC and PE plasmalogens, and short-chain TAGs were inversely associated. Caffeine, 7-dimethyluric, trigonelline, PE, and PC plasmalogens were the most positively associated with moderate alcohol intake, and both long- and short-chain TAGs and DAGs were inversely associated. Metabolites that were most positively associated with moderate to vigorous physical activity were PCs, PC plasmalogens, lysophosphatidylcholines, and CEs, whereas short-chain TAGs and DAGs were inversely associated. Proline betaine, biliverdin, C36:4 DAG, C22:CE 5, and long-chain TAGs were the most positively associated with never smoking, while trigonelline, caffeine, 5-acetylamino-6-amino-3-methyluracil, and 7-dimethyluric acid were inversely associated.

Healthy lifestyle metabolomic signature

A total of 101 metabolites associated with the self-reported healthy lifestyle score were selected from the elastic net regression (Figure 3). Selected metabolites were consistent with those identified from metabolome-wide analysis. For example, CEs (C22:5, C22:6, C16:0), PCs (C40:9, C38:6), and PC plasmalogens (C34:3, C36:3) were among the positively associated metabolites, and C18:0 SM, N2,N2-dimethylguanosine, creatine, N4-acetylcytidine, myristoleic acid, uric acid, L-valine, and shorter-chain saturated TAGs (C51:3, C53:3, C50:3, C52:2, C50:4) were among the inversely associated ones. These metabolites were also the most consistently

A

Top 40 metabolites most strongly associated with the healthy lifestyle score

Positive			Negative		
Metabolite	β (95% CI)	P value	Metabolite	β (95% CI)	P value
C22:6 CE	0.27 (0.25, 0.29)	2.6E-141	C52:2 TAG	-0.26 (-0.28, -0.24)	2.4E-129
C58:9 TAG	0.22 (0.20, 0.24)	5.8E-95	C50:3 TAG	-0.25 (-0.27, -0.23)	6.6E-119
C18:2 CE	0.21 (0.19, 0.23)	4.2E-82	C50:1 TAG	-0.24 (-0.26, -0.22)	1.5E-109
C40:9 PC	0.21 (0.19, 0.23)	4.2E-85	C50:2 TAG	-0.24 (-0.26, -0.22)	2.6E-111
C16:0 CE	0.20 (0.18, 0.22)	9.1E-78	C32:1 DAG	-0.24 (-0.26, -0.22)	6.1E-109
C38:6 PC	0.20 (0.18, 0.22)	3.6E-78	C34:1 DAG	-0.24 (-0.26, -0.22)	9.7E-107
C40:10 PC	0.18 (0.16, 0.20)	7.6E-59	C51:2 TAG	-0.23 (-0.25, -0.21)	8.6E-100
C34:2 PC plasmalogen	0.18 (0.16, 0.20)	1.7E-59	C52:1 TAG	-0.23 (-0.25, -0.21)	7.6E-103
C38:7 PC plasmalogen	0.18 (0.15, 0.20)	6.9E-57	C36:1 DAG	-0.23 (-0.25, -0.21)	1.3E-97
C34:3 PC plasmalogen	0.17 (0.15, 0.19)	2.9E-60	C36:2 DAG	-0.23 (-0.25, -0.21)	2.5E-99
C20:5 CE	0.15 (0.12, 0.17)	4.2E-42	C53:2 TAG	-0.23 (-0.25, -0.21)	2.9E-101
C22:5 CE	0.15 (0.13, 0.17)	1.1E-46	C54:1 TAG	-0.22 (-0.24, -0.20)	8.7E-93
C22:0 LPE	0.15 (0.12, 0.17)	1.7E-37	C34:2 DAG	-0.22 (-0.24, -0.20)	3.8E-93
C40:7 PE plasmalogen	0.14 (0.11, 0.16)	1.1E-34	C51:1 TAG	-0.22 (-0.24, -0.20)	4.9E-95
Asparagine	0.13 (0.11, 0.15)	2.3E-28	C48:2 TAG	-0.21 (-0.23, -0.19)	2.6E-88
C56:8 TAG	0.13 (0.11, 0.15)	3.4E-33	C52:3 TAG	-0.21 (-0.23, -0.19)	2.3E-82
C36:3 PC plasmalogen	0.13 (0.11, 0.15)	1.4E-34	C54:2 TAG	-0.21 (-0.24, -0.19)	5.9E-86
C38:7 PE plasmalogen	0.13 (0.11, 0.15)	6.8E-32	C38:3 PC	-0.21 (-0.23, -0.18)	1.4E-76
C22:6 LPE	0.13 (0.10, 0.15)	5.8E-30	C49:3 TAG	-0.21 (-0.23, -0.19)	1.7E-83
C18:2 LPC	0.12 (0.10, 0.14)	2.6E-141	C48:1 TAG	-0.20 (-0.22, -0.18)	2.4E-129

B

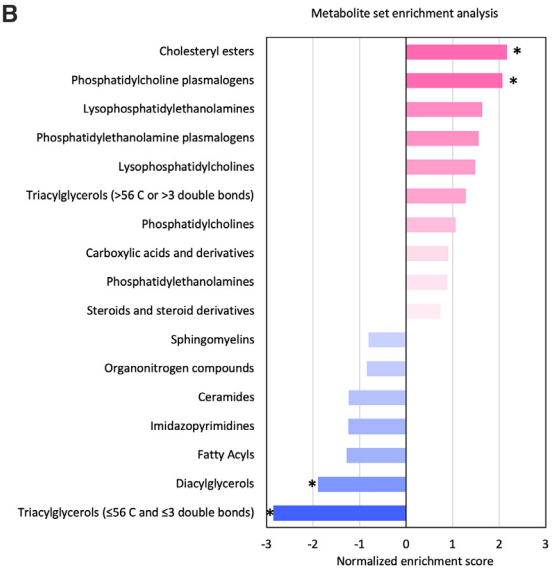


Figure 2. Metabolome-wide associations for the healthy lifestyle adherence

(A) Top 40 positive and inverse associations between individual metabolites and healthy lifestyle adherence. Results are from multivariable linear regressions adjusted for age, fasting status, ethnicity, multivitamin use, diabetes, hypertension, antihypertensive medication use, hypercholesterolemia, lipid-lowering medication use, total energy intake, study cohorts, original substudies, and the case/control status within the original substudy. p values are corrected for multiple testing using a Bonferroni adjustment.

(B) Normalized enrichment score of predefined metabolite sets with adherence to a healthy lifestyle. Results are from metabolite set enrichment analysis (MSEA) based on individual multivariable-adjusted linear regression β coefficients. p values are corrected for multiple testing using a false discovery rate (FDR) adjustment. β, standardized beta coefficient; CI, confidence interval; CE, cholesteryl esters; PC, phosphatidylcholine; TAG, triacylglycerols; PE, phosphatidylethanolamine. *$p < 0.05$.

associated with individual healthy lifestyle factors (Figure 3, second column). The healthy lifestyle metabolomic signature correlated with the self-reported healthy lifestyle score in both the training (Pearson r = 0.46, $r^2 = 0.21$; $p < 0.001$) and testing sets (Pearson r = 0.45, $r^2 = 0.20$; $p < 0.001$) and across cohorts (NHS: Pearson r = 0.43, $r^2 = 0.19$; NHSII: r = 0.49, $r^2 = 0.24$; HPFS: r = 0.47, $r^2 = 0.22$; all $p < 0.001$) (Figures 4A–4C). Individual healthy lifestyle factors correlated with the healthy lifestyle metabolomic signature. Having a BMI between 18.5 and 24.9 kg/m^2 was the most strongly correlated (point-biserial r = 0.43, $p < 0.001$), followed by a healthier diet (point-biserial r = 0.27; $p < 0.001$), moderate to vigorous physical activity (point-biserial r = 0.20; $p < 0.001$), low-moderate alcohol intake (point-biserial r = 0.14; $p < 0.001$), and never smoking (point-biserial r = 0.04; $p < 0.001$). In multivariable-adjusted logistic regressions, the healthy lifestyle metabolomic signature was significantly associated with each individual lifestyle factor, independently of other factors. The odds ratios per 1-standard deviation (SD) increment in the healthy lifestyle metabolomic signature and the 95% confidence interval (CI) for the individual lifestyle factors were 9.39 (8.40, 10.49; "normal" BMI), 4.20 (3.79, 4.65; healthy diet), 1.89 (1.69, 2.12; moderate alcohol), 1.85 (1.66, 2.05; never smoking), and 1.42 (1.29, 1.56; active living).

Prospective associations with mortality and longevity

Over up to 28 years of follow-up, we documented 4,343 deaths (3,126 in NHS, 155 in NHSII, and 1,062 in HPFS), including 894 cardiovascular disease (CVD) and 1,108 cancer deaths. As expected, a higher self-reported adherence to a healthy lifestyle was associated with a lower risk of all-cause mortality (hazard ratio [HR] per 1-SD increment = 0.86, 95% CI: 0.81, 0.92), CVD mortality (HR per 1-SD increment = 0.79, 95% CI: 0.67, 0.94), and cancer mortality (HR per 1-SD increment = 0.89,

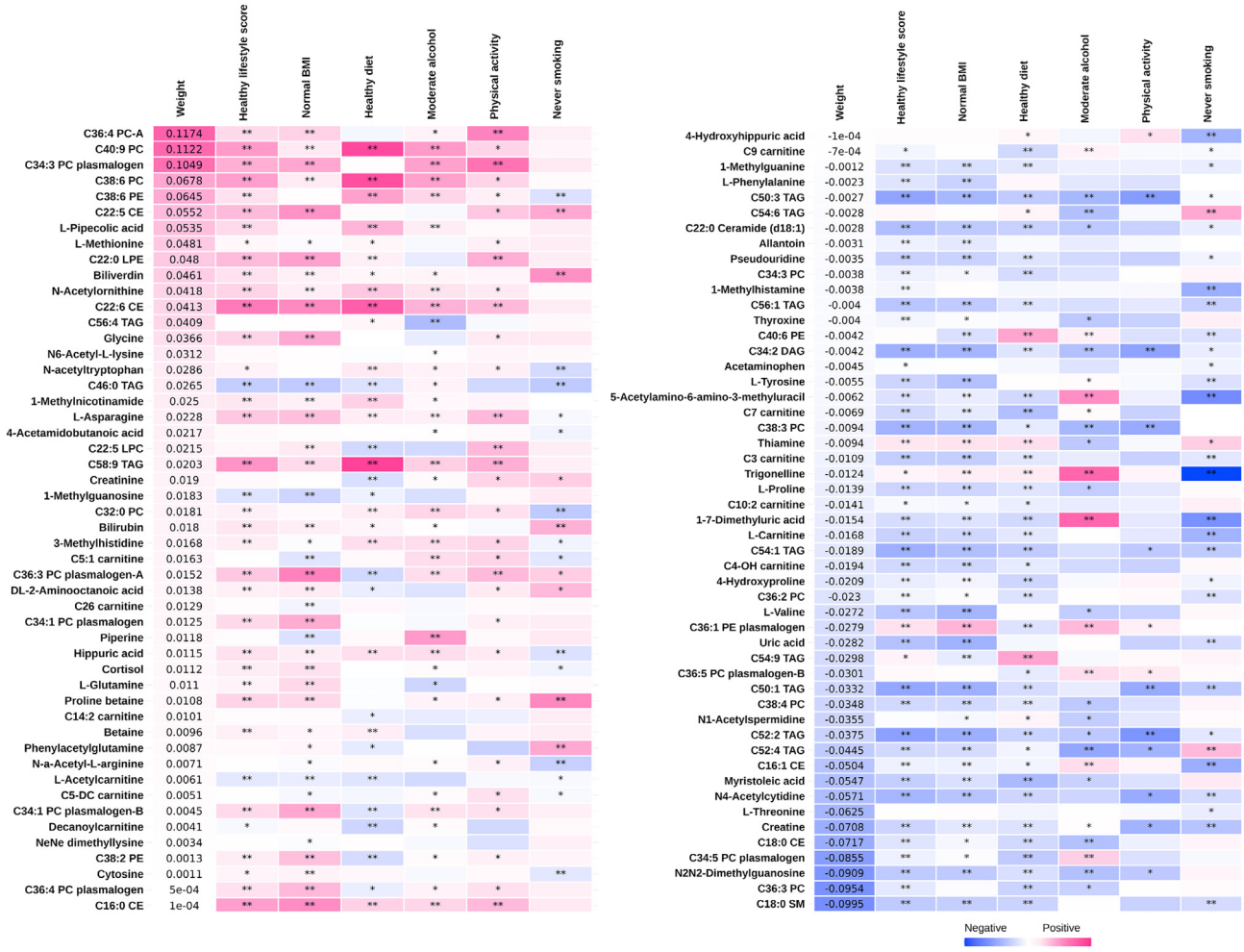


Figure 3. Associations between individual metabolites and healthy lifestyle adherence and its components

This heatmap shows the 101 metabolites retained in the healthy lifestyle signature and their weight from regularized elastic net regressions, β coefficient from multivariable-adjusted linear regression (self-reported healthy lifestyle adherence), and odds ratios (ORs) from multivariable-adjusted logistic regressions (alcohol, BMI, diet, physical activity, smoking). Results are adjusted for age, fasting status, ethnicity, multivitamin use, diabetes, hypertension, antihypertensive medication use, hypercholesterolemia, lipid-lowering medication use, total energy intake, study cohorts, original substudies, and the case/control status within the original substudy. Results for each healthy lifestyle factor are further adjusted for the other healthy lifestyle factors. p values are corrected for multiple testing using a Bonferroni adjustment. Dark pink indicates a strong positive association and dark blue a strong negative association. *$p < 0.05$ and **$p < 0.001$.

95% CI: 0.80, 1.00) (Table 2). The healthy lifestyle metabolomic signature was also associated with lower risk of all-cause mortality (HR per 1-SD increment = 0.83, 95% CI: 0.78, 0.89), CVD mortality (HR per 1-SD increment = 0.81, 95% CI: 0.68, 0.96), and cancer mortality (HR per 1-SD increment = 0.83, 95% CI: 0.74, 0.92) (Table 2). The associations with total and cancer mortality, but not CVD mortality, remained significant after further adjustment for self-reported healthy lifestyle score. In mediation analysis, the healthy lifestyle metabolomic signature attenuated the association between the self-reported healthy lifestyle score and all-cause mortality and explained 38.0% (95% CI: 21.1, 58.4; $p < 0.0001$) of the association.

Longevity ≥85 years of age was achieved by 2,616 participants (n = 2,396 deaths before the age of 85 years). Higher self-reported adherence to a healthy lifestyle was associated with greater likelihood of achieving longevity (HR per 1-SD

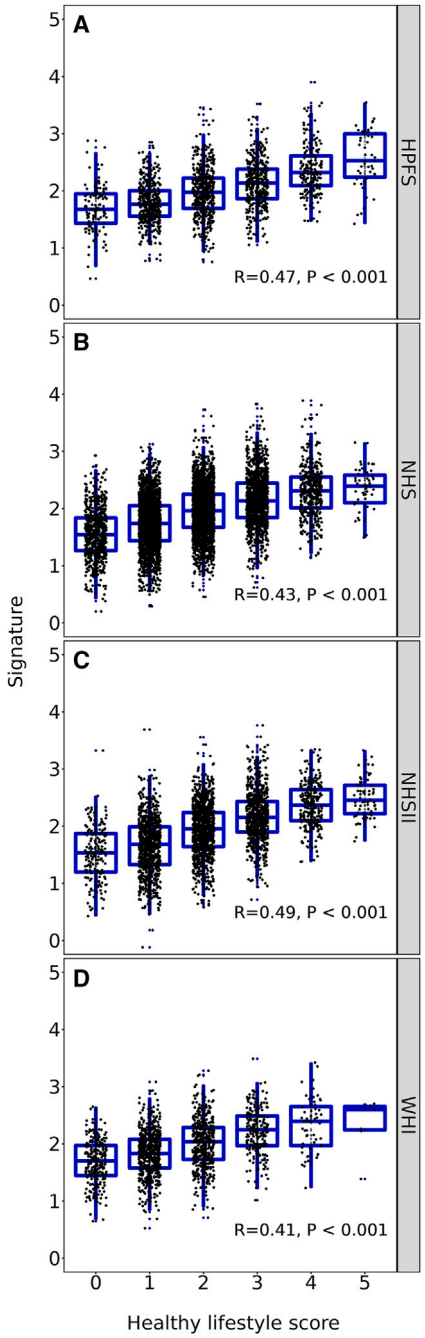


Figure 4. Correlation between the healthy lifestyle metabolomic signature and the self-reported score by cohorts

(A) Correlation in the HPFS.

(B) Correlation in the NHS.

(C) Correlation in the NHSII.

(D) Correlation in the WHI study.

For each of the number of healthy factors (0, 1, 2, 3, 4, 5; x axis), boxplots represent the interquartile range and median of the corresponding number of factors estimated by the metabolomic signature (y axis). R is the Pearson correlation coefficient.

Table 2. Associations of the healthy lifestyle adherence and metabolomic signature with total and cause-specific mortality and longevity in NHS, NHSII, and HPFS

	Healthy lifestyle adherence per 1-SD increment		Metabolomic signature per 1-SD increment	
	HR (95% CI)	p	HR (95% CI)	p
All-cause mortality n deaths = 4,343/11,487				
Model 1	0.84 (0.79, 0.90)	<0.001	0.81 (0.76, 0.86)	<0.001
Model 2	0.86 (0.80, 0.92)	<0.001	0.83 (0.78, 0.89)	<0.001
Model 3	0.91 (0.85, 0.98)	0.011	0.86 (0.80, 0.92)	<0.001
CVD mortality n deaths = 894/11,487				
Model 1	0.77 (0.65, 0.90)	0.001	0.74 (0.63, 0.88)	0.001
Model 2	0.79 (0.67, 0.94)	0.008	0.81 (0.68, 0.96)	0.015
Model 3	0.83 (0.70, 0.99)	0.04	0.85 (0.71, 1.02)	0.086
Cancer mortality n deaths = 1,108/11,487				
Model 1	0.89 (0.80, 0.99)	0.031	0.82 (0.74, 0.92)	<0.001
Model 2	0.89 (0.80, 1.00)	0.047	0.82 (0.74, 0.92)	0.001
Model 3	0.96 (0.85, 1.08)	0.484	0.84 (0.74, 0.95)	0.005
Longevity ≥85 years n longevity = 2,616/11,487 n death <85 years = 2,396/11,487				
Model 1	1.21 (1.12, 1.30)	<0.001	1.29 (1.20, 1.39)	<0.001
Model 2	1.17 (1.09, 1.27)	<0.001	1.25 (1.16, 1.35)	<0.001
Model 3	1.09 (1.00, 1.18)	0.055	1.22 (1.12, 1.32)	<0.001

HRs are per 1-SD increment. Model 1 was adjusted for fasting status and stratified by baseline age, study cohorts, original substudies, and case-control status within the original substudy. Model 2 was further adjusted for race/ethnicity, multivitamin use, diabetes, hypercholesterolemia, lipid-lowering medication use, hypertension, antihypertensive medication use, and total caloric intake. Model 3 further included adjustment for the self-reported healthy lifestyle adherence to examine association independence. SD, standard deviation; HR, hazard ratio; CI, confidence interval.

increment = 1.17, 95% CI: 1.09, 1.27). A higher healthy lifestyle metabolomic score was also associated with a greater likelihood of achieving longevity (HR per 1-SD increment = 1.25, 95% CI: 1.16, 1.35) and remained after mutual adjustment for the self-reported healthy lifestyle score (HR per 1-SD increment = 1.22, 95% CI: 1.12, 1.32) (Table 2). The healthy lifestyle metabolomic signature explained 48.6% (95% CI: 25.4–72.4; $p < 0.0001$) of the association between the self-reported healthy lifestyle score and longevity.

External replication

In the WHI cohort, we replicated 213 metabolite associations, of which 34 were positively associated with the self-reported healthy lifestyle score and 55 were inversely associated (Bonferroni-adjusted $p < 0.05$) (Table S10). Highly concordant with the results we found in the NHS/NHSII/HPFS, PCs (C40:10, C38:6), CEs (namely C22:6, C20:5, C16:0, and C22:5), and long-chain unsaturated TAGs (C58:9, C56:8) were the most positively associated with a healthy lifestyle, whereas higher-saturation and shorter-chain TAGs (including C52:2, C54:1, C54:2, C52:3, C51:3, C52:1, and C51:1) and DAGs (including C36:1, C34:1, C36:2, and C34:2) were inversely associated (all Bonferroni-adjusted $p < 0.05$). The healthy lifestyle metabolomic signature also correlated with the combined self-reported healthy lifestyle score (Figure 4D; Pearson $r = 0.41$, $p < 0.001$). Results of the associations between the healthy lifestyle metabolomic signature and all-cause mortality (HR per 1-SD increment = 0.82, 95%

Table 3. Associations of the healthy lifestyle adherence and metabolomic signature with total and cause-specific mortality and longevity in WHI

	Healthy lifestyle adherence per 1-SD increment		Metabolomic signature per 1-SD increment	
	HR (95% CI)	p value	HR (95% CI)	p value
All-cause mortality N deaths = 679/1,569				
Model 1	0.83 (0.76, 0.90)	<0.001	0.79 (0.73, 0.86)	<0.001
Model 2	0.85 (0.78, 0.92)	<0.001	0.82 (0.76, 0.89)	<0.001
Model 3	0.90 (0.82, 0.98)	0.017	0.86 (0.78, 0.94)	<0.001
Longevity ≥85 years N longevity = 622/1,569 N deaths <85 years = 169/1,569				
Model 1	1.39 (1.17, 1.66)	<0.001	1.43 (1.22, 1.66)	<0.001
Model 2	1.37 (1.14, 1.64)	<0.001	1.39 (1.18, 1.62)	<0.001
Model 3	1.21 (0.99, 1.48)	0.057	1.29 (1.08, 1.54)	0.004

HRs are per 1-SD increment. Model 1 was stratified by baseline age. Model 2 was further adjusted for race/ethnicity, multivitamin use, diabetes, hypertension, antihypertensive medication use, and total caloric intake. Model 3 further included adjustment for the self-reported healthy lifestyle adherence to examine association independence. SD, standard deviation; HR, hazard ratio; CI, confidence interval.

CI: 0.76, 0.89) and longevity (HR per 1-SD increment = 1.39, 95% CI: 1.18, 1.62) were also highly consistent with those observed in the NHS/NHSII/HPFS (Table 3).

DISCUSSION

Based on data from four large US prospective cohorts of 13,056 participants, this study identifies a metabolomic signature that measures adherence to a healthy lifestyle and is among the first studies to show prediction of total and cause-specific mortality and longevity. The metabolomic signature was mostly reflective of lipid metabolism pathways involving PC, CE, TAG, and DAG metabolite groups. Among lifestyle components, the metabolomic signature was most strongly associated with having a "normal" BMI and a healthy diet. We also identified key metabolites associated with each lifestyle component individually. The healthy lifestyle metabolomic signature's ability to predict mortality risk and longevity may be useful in complementing traditional questionnaires and personalizing lifestyle-based interventions.

Observational and intervention studies have uncovered metabolites associated with individual lifestyle factors including BMI,[5,17–20] diet quality,[21] alcohol intake,[8,9] physical activity,[7] and smoking[10,11] in adults, but very few examined associations with these factors combined and covering a broad range of metabolites.[12,14,15] In our study, the healthy lifestyle signature was composed of 101 metabolites (101/243; 41.5% of measured metabolites) selected by elastic net regression; specifically, 58 were positively associated and 129 were inversely associated. It is not surprising that several plasma metabolites may be useful in assessing lifestyle status given the important effect of lifestyle behaviors on the metabolism.[22] A study that measured 1,251 serum metabolites in 491 well-phenotyped healthy individuals reported that 48.9% of the metabolite-type enrichment was predicted by diet, 47.5% by clinical data including anthropometrics, and 2% by other lifestyle factors such as smoking and exercise (relative to full model predictive power).[22] This is in accordance with our findings highlighting the strongest associations of our healthy lifestyle metabolomic signature with the BMI and diet lifestyle factors.

Our metabolome-wide analyses underpinned pathways associated with a healthy lifestyle that are highly reflective of lipid metabolism pathways. Indeed, the MSEA

revealed CEs, mainly of PUFAs, and PCs as the most enriched metabolite sets positively associated with a healthy lifestyle. CEs serve as a mean for the storage and transportation of cholesterol and other lipids in the blood[23] and were shown to be reflective of dietary fat intake.[24] PCs are naturally found in the body but also in foods such as eggs, fatty fish, and soybeans.[25] They are well known for their essential role in cell membranes and membrane signaling.[26] In the lifestyle metabolomic signature, C36:4 PC-A, C40:9 PC, and C34:3 PC plasmalogen were the most influential metabolites that showed positive associations with a healthier lifestyle. C40:9 PC has been associated with fish intake[27] and diet quality.[28] In line with our results, this metabolite was most strongly associated with a healthier diet among lifestyle components. Previous studies reported higher C36:4 PC-A and C34:3 PC plasmalogen with greater physical activity level,[29] lower BMI,[30] and higher alcohol consumption,[8] which is consistent with our findings showing associations of these metabolites with meeting the recommendations for physical activity and alcohol intake and with having a BMI between 18.5 and 24.9 kg/m^2. On the other hand, we found that higher levels of several shorter-chain and saturated TAGs and DAGs were indicative of a poorer lifestyle. Triglyceride-rich lipoproteins are well recognized for their atherogenic effect,[31] and intervention studies support a beneficial role of a healthy lifestyle in reducing circulating TAGs.[32] In the healthy lifestyle metabolomic signature, C18:0 SM was the most influential metabolite with an inverse association. Sphingolipids are the most commonly occurring types of lipids present in circulating low-density lipoprotein and were related to an increased risk of metabolic diseases.[33] Palau-Rodriguez et al. reported lower SMs (d18:0/22:0 and d18:0/20:0, d16:0/22:0) in the high (>10%) compared to the low (<10%) weight loss group following a 12-month lifestyle intervention (hypocaloric Mediterranean diet and increased physical activity) in 27 women with obesity.[34] Another study found an association between higher SM (d18:0/20:0, d16:0/22:0) and higher BMI in a nonalcoholic fatty liver disease case-subcohort study (n = 356).[18] As for diet, Shah et al. found an association between higher SMs and higher red meat, fish, and chicken and lower dark green vegetables, fruits, and whole grains in the Coronary Artery Risk Development in Young Adults (CARDIA) study (n = 2,259 White and Black adults).[35] As corroborated by our results, the inverse association of SM C18:0 with a healthy lifestyle appears to be driven by a BMI between 18.5 and 24.9 kg/m^2, a healthier diet, and possibly never-smoking status.

Although several lipid species were retained in the signature, other species such as amino acids and metabolites of the purine metabolism were also significantly associated with a healthy lifestyle. In previous studies, lower circulating glycine[5,19] and asparagine[5,20,36] and higher tyrosine,[5,19,20] valine,[5,19] and uric acid[5,17,37] have been commonly associated with higher BMI or obesity. In the current study, and consistent with previous studies, uric acid and amino acids including glycine, asparagine, tyrosine, and valine were comparably associated with BMI and were selected in the lifestyle metabolomic signature. Higher circulating glycine, trigonelline, asparagine, hippurate, and glutamine and lower valine, isoleucine, and leucine were also observed with lower intakes of red meat, chicken, and diet drinks and higher intakes of dark green vegetables, fruits, and whole grains in the CARDIA study.[35] In our study, glycine, trigonelline, asparagine, hippuric acid, glutamine, and valine were retained in the healthy lifestyle signature; however, only asparagine, hippuric acid, and trigonelline were associated with a healthier diet. Trigonelline, an alkaloid abundant in coffee beans and found in other plants, is thought to have neuroprotective and antidiabetic effects given its antioxidant properties.[38] While positively associated with overall adherence to a healthy lifestyle, healthier diet, and low-moderate alcohol intake, trigonelline (and caffeine) was also inversely associated

with never smoking in our study. This may be explained by faster caffeine metabolism and higher coffee intake among smokers compared to nonsmokers.[39,40]

The impact of health behaviors on premature mortality risk has been well studied. A healthy lifestyle was associated with 53%–82% estimated lower mortality risk in different populations worldwide including the US,[3,41–43] Japan,[44] China,[45] Germany,[46,47] Denmark,[47] Norway,[47] and the UK.[2] Specifically, in the NHS (1980–2014; n = 78,865) and HPFS (1986–2014, n = 44,354) cohorts, Li et al. showed that individuals adhering to 5 healthy lifestyle factors had a lower risk of total mortality (HR = 0.26, 95% CI: 0.22–0.31) and cause-specific mortality (HR for CVD = 0.18, 95% CI: 0.12–0.26; HR for cancer = 0.35, 95% CI: 0.27–45) compared to nonadherents (zero healthy lifestyle factors).[3] In the current metabolomic study, we found an association between the healthy lifestyle metabolomic signature with lower total and cause-specific mortality. Interestingly, the association with lower mortality was stronger for the healthy lifestyle metabolomic signature compared to that of the self-reported healthy lifestyle score. Indeed, the metabolomic signature explained 38.0% of the association between the self-reported healthy lifestyle score and mortality, pointing to unique biological pathways captured by metabolomics. Specific lifestyle behaviors including hypercaloric nutrition and sedentariness are known to alter the metabolism such that it may accelerate aging.[48] Consistent with the literature and with our mortality results, we found an association of the healthy lifestyle metabolomic signature with longevity, and the signature explained 48.6% of the association between self-reported healthy lifestyle score and longevity.

Our study contributes new insights to the limited evidence available on metabolites associated with a combined healthy lifestyle and their associations with mortality risk and longevity. Our findings were highly reproducible among 4 cohorts including men and women from across the US with varying lifestyle habits. The signature was derived using elastic net regularization models with training-test validation sets and using a leave-one-out approach for the training set to reduce bias and overfitting and was replicated in a separate cohort (WHI). Metabolomic profiling was conducted using systematic and rigorous multiplatform liquid chromatography mass spectrometry methods in all cohorts including the replication dataset. The wide range of metabolites measured will facilitate future study comparison. Lastly, the large sample size, long follow-up period, and ascertained deaths support the robustness of our results.

Overall, our findings suggest that greater adherence to a healthy lifestyle may lead to alterations in the metabolome that are associated with lower premature mortality risk and higher likelihood of longevity. We identified a metabolomic signature associated with a combined healthy lifestyle in US adults that is strongly reflective of lipid metabolism pathways. We found that those with a higher multimetabolite score had a lower risk of total and cause-specific mortality and a greater likelihood of living longer. These results shed new light on key metabolites and metabolic pathways associated with lifestyle and hold promising clinical relevance for targeted interventions aimed at reducing the risk of premature mortality and promoting longevity through metabolic health improvement.

Limitation of the study

Our findings should be interpreted in the context of the study limitations. First, association between the self-reported healthy lifestyle score and its metabolomic signature may be bidirectional given the cross-sectional nature of analyses. Second, blood was analyzed in case-control endpoint batches, potentially introducing variability in measurements; nonetheless, we accounted for batch effect in all models

by adjusting or stratifying by endpoints. We also adjusted models for case-control status, and our stratified analysis showed consistent results in both cases and controls. Third, the results of mediation analyses rely on a causal assumption and may have been affected by unmeasured confounding factors, such as genetics. Fourth, this study was conducted among health professionals, who may be healthier than the general population. Lastly, this study was limited to plasma metabolites of known identity, potentially biasing findings toward available metabolites and classes. Future work should include untargeted metabolomic profiling to address this limitation. Similar studies should be conducted in populations of various ethnicities and socio-economic statuses. Studies examining the changes in metabolomic profile and its association with mortality and longevity are also warranted.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Blood collection and metabolite profiling
  - Assessment of lifestyle factors
  - Ascertainment of deaths and longevity
  - Assessment of covariates
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.medj.2024.01.010.

ACKNOWLEDGMENTS

This work is supported by the research grant R21 AG070375 from the National Institutes of Health to M.G.-F. The NHS, NHSII, HPFS, and WHI are supported by grants from the National Institutes of Health (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 DK112940, R01 DK119268, R01 DK120870, P30 DK046200, HHSN268201300008C, HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, and HHSN268201600004C). A.-J.T. is supported by the Canadian Institutes of Health Research (CIHR) Fellowship Award. F.W. is supported by the American Heart Association Postdoctoral Fellowship (grant number: 897161). M.G.-F. is supported by Novo Nordisk Foundation grant NNF18CC0034900. C.W. is supported by the SciLifeLab and Wallenberg Data Driven Life Science Program (grant KAW 2020.0239). 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. The authors would like to acknowledge the contribution to this study 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 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, and Wyoming.

AUTHOR CONTRIBUTIONS

A.-J.T., M.G.-F., and F.B.H. designed the research and had unrestricted access to all data. A.-J.T. and F.W. conducted analyses. A.H.E. and C.B.C. participated in acquisition of data. L.L. and O.Z. provided statistical expertise. A.-J.T. prepared tables and figures and wrote the manuscript. All authors contributed to the interpretation of the results and critical revision of the manuscript for important intellectual content. All authors approved the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 9, 2023
Revised: November 9, 2023
Accepted: January 18, 2024
Published: February 15, 2024

REFERENCES

1. Zhang, Y.-B., Pan, X.-F., Chen, J., Cao, A., Xia, L., Zhang, Y., Wang, J., Li, H., Liu, G., and Pan, A. (2021). Combined lifestyle factors, all-cause mortality and cardiovascular disease: a systematic review and meta-analysis of prospective cohort studies. J. Epidemiol. Community Health *75*, 92–99. https://doi.org/10.1136/jech-2020-214050.

2. Han, H., Cao, Y., Feng, C., Zheng, Y., Dhana, K., Zhu, S., Shang, C., Yuan, C., and Zong, G. (2022). Association of a Healthy Lifestyle With All-Cause and Cause-Specific Mortality Among Individuals With Type 2 Diabetes: A Prospective Study in UK Biobank. Diabetes Care *45*, 319–329. https://doi.org/10.2337/dc21-1512.

3. Li, Y., Pan, A., Wang, D.D., Liu, X., Dhana, K., Franco, O.H., Kaptoge, S., Di Angelantonio, E., Stampfer, M., Willett, W.C., and Hu, F.B. (2018). Impact of Healthy Lifestyle Factors on Life Expectancies in the US Population. Circulation *138*, 345–355. https://doi.org/10.1161/CIRCULATIONAHA.117.032047.

4. Griffin, J.L., Atherton, H., Shockcor, J., and Atzori, L. (2011). Metabolomics as a tool for cardiac research. Nat. Rev. Cardiol. *8*, 630–643. https://doi.org/10.1038/nrcardio.2011.138.

5. Pan, X.-F., Chen, Z.-Z., Wang, T.J., Shu, X., Cai, H., Cai, Q., Clish, C.B., Shi, X., Zheng, W., Gerszten, R.E., et al. (2022). Plasma metabolomic signatures of obesity and risk of type 2 diabetes. Obesity *30*, 2294–2306. https://doi.org/10.1002/oby.23549.

6. Guasch-Ferré, M., Bhupathiraju, S.N., and Hu, F.B. (2018). Use of Metabolomics in Improving Assessment of Dietary Intake. Clin. Chem. *64*, 82–98. https://doi.org/10.1373/clinchem.2017.272344.

7. Kelly, R.S., Kelly, M.P., and Kelly, P. (2020). Metabolomics, physical activity, exercise and health: A review of the current evidence. Biochim. Biophys. Acta, Mol. Basis Dis. *1866*, 165936. https://doi.org/10.1016/j.bbadis.2020.165936.

8. van Roekel, E.H., Trijsburg, L., Assi, N., Carayol, M., Achaintre, D., Murphy, N., Rinaldi, S., Schmidt, J.A., Stepien, M., Kaaks, R., et al. (2018). Circulating Metabolites Associated with Alcohol Intake in the European Prospective Investigation into Cancer and Nutrition Cohort. Nutrients *10*, 654. https://doi.org/10.3390/nu10050654.

9. Dorgan, J.F., Jung, S., Dallal, C.M., Zhan, M., Stennett, C.A., Zhang, Y., Eckert, R.L., Snetselaar, L.G., and Van Horn, L. (2020). Alcohol consumption and serum metabolite concentrations in young women. Cancer Causes Control. *31*, 113–126. https://doi.org/10.1007/s10552-019-01256-1.

10. Gu, F., Derkach, A., Freedman, N.D., Landi, M.T., Albanes, D., Weinstein, S.J., Mondul, A.M., Matthews, C.E., Guertin, K.A., Xiao, Q., et al. (2016). Cigarette smoking behaviour and blood metabolomics. Int. J. Epidemiol. *45*, 1421–1432. https://doi.org/10.1093/ije/dyv330.

11. Zhang, R., Sun, X., Huang, Z., Pan, Y., Westbrook, A., Li, S., Bazzano, L., Chen, W., He, J., Kelly, T., and Li, C. (2022). Examination of serum metabolome altered by cigarette smoking identifies novel metabolites mediating smoking-BMI association. Obesity *30*, 943–952. https://doi.org/10.1002/oby.23386.

12. Assi, N., Gunter, M.J., Thomas, D.C., Leitzmann, M., Stepien, M., Chajès, V., Philip, T., Vineis, P., Bamia, C., Boutron-Ruault, M.-C., et al. (2018). Metabolic signature of healthy lifestyle and its relation with risk of hepatocellular carcinoma in a large European cohort. Am. J. Clin. Nutr. *108*, 117–126. https://doi.org/10.1093/ajcn/nqy074.

13. Fu, Z., Liu, Q., Liang, J., Weng, Z., Li, W., Xu, J., Zhang, X., Xu, C., and Gu, A. (2023). Association between NMR metabolomic signatures of healthy lifestyle and incident coronary artery disease. Eur. J. Prev. Cardiol. *30*, 243–253. https://doi.org/10.1093/eurjpc/zwac252.

14. Rothwell, J.A., Murphy, N., Bešević, J., Kliemann, N., Jenab, M., Ferrari, P., Achaintre, D., Gicquiau, A., Vozar, B., Scalbert, A., et al. (2022). Metabolic Signatures of Healthy Lifestyle Patterns and Colorectal Cancer Risk in a European Cohort. Clin. Gastroenterol. Hepatol. *20*, e1061–e1082. https://doi.org/10.1016/j.cgh.2020.11.045.

15. Delgado-Velandia, M., Gonzalez-Marrachelli, V., Domingo-Relloso, A., Galvez-Fernandez, M., Grau-Perez, M., Olmedo, P., Galan, I., Rodriguez-Artalejo, F., Amigo, N., Briongos-Figuero, L., et al. (2022). Healthy lifestyle, metabolomics and incident type 2 diabetes in a population-based cohort from Spain. Int. J. Behav. Nutr. Phys. Activ. *19*, 8. https://doi.org/10.1186/s12966-021-01219-3.

16. Kaspy, M.S., Semnani-Azad, Z., Malik, V.S., Jenkins, D.J.A., and Hanley, A.J. (2022). Metabolomic profile of combined healthy lifestyle behaviours in humans: A systematic

review. Proteomics *22*, e2100388. https://doi.org/10.1002/pmic.202100388.

17. McClain, K.M., Friedenreich, C.M., Matthews, C.E., Sampson, J.N., Check, D.P., Brenner, D.R., Courneya, K.S., Murphy, R.A., and Moore, S.C. (2022). Body Composition and Metabolomics in the Alberta Physical Activity and Breast Cancer Prevention Trial. J. Nutr. *152*, 419–428. https://doi.org/10.1093/jn/nxab388.

18. Pang, Y., Kartsonaki, C., Lv, J., Millwood, I.Y., Fairhurst-Hunter, Z., Turnbull, I., Bragg, F., Hill, M.R., Yu, C., Guo, Y., et al. (2022). Adiposity, metabolomic biomarkers, and risk of nonalcoholic fatty liver disease: a case-cohort study. Am. J. Clin. Nutr. *115*, 799–810. https://doi.org/10.1093/ajcn/nqab392.

19. Tan, H.C., Hsu, J.W., Tai, E.S., Chacko, S., Wu, V., Lee, C.F., Kovalik, J.-P., and Jahoor, F. (2022). De Novo Glycine Synthesis Is Reduced in Adults With Morbid Obesity and Increases Following Bariatric Surgery. Front. Endocrinol. *13*, 900343. https://doi.org/10.3389/fendo.2022.900343.

20. Bagheri, M., Djazayery, A., Farzadfar, F., Qi, L., Yekaninejad, M.S., Aslibekyan, S., Chamari, M., Hassani, H., Koletzko, B., and Uhl, O. (2019). Plasma metabolomic profiling of amino acids and polar lipids in Iranian obese adults. Lipids Health Dis. *18*, 94. https://doi.org/10.1186/s12944-019-1037-0.

21. Li, J., Guasch-Ferré, M., Chung, W., Ruiz-Canela, M., Toledo, E., Corella, D., Bhupathiraju, S.N., Tobias, D.K., Tabung, F.K., Hu, J., et al. (2020). The Mediterranean diet, plasma metabolome, and cardiovascular disease risk. Eur. Heart J. *41*, 2645–2656. https://doi.org/10.1093/eurheartj/ehaa209.

22. Bar, N., Korem, T., Weissbrod, O., Zeevi, D., Rothschild, D., Leviatan, S., Kosower, N., Lotan-Pompan, M., Weinberger, A., Le Roy, C.I., et al. (2020). A reference map of potential determinants for the human serum metabolome. Nature *588*, 135–140. https://doi.org/10.1038/s41586-020-2896-2.

23. Rader, D.J. (2006). Molecular regulation of HDL metabolism and function: implications for novel therapies. J. Clin. Invest. *116*, 3090–3100. https://doi.org/10.1172/JCI30163.

24. Rosqvist, F., Fridén, M., Vessby, J., Rorsman, F., Lind, L., and Risérus, U. (2022). Circulating fatty acids from high-throughput metabolomics platforms as potential biomarkers of dietary fatty acids. Clin. Nutr. *41*, 2637–2643. https://doi.org/10.1016/j.clnu.2022.10.005.

25. Wiedeman, A.M., Barr, S.I., Green, T.J., Xu, Z., Innis, S.M., and Kitts, D.D. (2018). Dietary choline intake: current state of knowledge across the life cycle. Nutrients *10*, 1513. https://doi.org/10.3390/nu10101513.

26. Furse, S., and de Kroon, A.I.P.M. (2015). Phosphatidylcholine's functions beyond that of a membrane brick. Mol. Membr. Biol. *32*, 117–119. https://doi.org/10.3109/09687688.2015.1066894.

27. Mazzilli, K.M., McClain, K.M., Lipworth, L., Playdon, M.C., Sampson, J.N., Clish, C.B., Gerszten, R.E., Freedman, N.D., and Moore, S.C. (2020). Identification of 102 Correlations between Serum Metabolites and Habitual Diet in a Metabolomics Study of the Prostate, Lung, Colorectal, and Ovarian Cancer Trial. J. Nutr. *150*, 694–703. https://doi.org/10.1093/jn/nxz300.

28. Bagheri, M., Willett, W., Townsend, M.K., Kraft, P., Ivey, K.L., Rimm, E.B., Wilson, K.M., Costenbader, K.H., Karlson, E.W., Poole, E.M., et al. (2020). A lipid-related metabolomic pattern of diet quality. Am. J. Clin. Nutr. *112*, 1613–1630. https://doi.org/10.1093/ajcn/nqaa242.

29. Ding, M., Zeleznik, O.A., Guasch-Ferré, M., Hu, J., Lasky-Su, J., Lee, I.-M., Jackson, R.D., Shadyab, A.H., LaMonte, M.J., Clish, C., et al. (2019). Metabolome-Wide Association Study of the Relationship Between Habitual Physical Activity and Plasma Metabolite Levels. Am. J. Epidemiol. *188*, 1932–1943. https://doi.org/10.1093/aje/kwz171.

30. Bellot, P.E.N.R., Moia, M.N., Reis, B.Z., Pedrosa, L.F.C., Tasic, L., Barbosa, F., and Sena-Evangelista, K.C.M. (2023). Are Phosphatidylcholine and Lysophosphatidylcholine Body Levels Potentially Reliable Biomarkers in Obesity? A Review of Human Studies. Mol. Nutr. Food Res. *67*, e2200568. https://doi.org/10.1002/mnfr.202200568.

31. Borén, J., Taskinen, M.-R., Björnson, E., and Packard, C.J. (2022). Metabolism of triglyceride-rich lipoproteins in health and dyslipidaemia. Nat. Rev. Cardiol. *19*, 577–592. https://doi.org/10.1038/s41569-022-00676-y.

32. Fernández-García, J.C., Martínez-Sánchez, M.A., Bernal-López, M.R., Muñoz-Garach, A., Martínez-González, M.A., Fitó, M., Salas-Salvadó, J., Tinahones, F.J., and Ramos-Molina, B. (2020). Effect of a lifestyle intervention program with energy-restricted Mediterranean diet and exercise on the serum polyamine metabolome in individuals at high cardiovascular disease risk: a randomized clinical trial. Am. J. Clin. Nutr. *111*, 975–982. https://doi.org/10.1093/ajcn/nqaa064.

33. Meikle, P.J., and Summers, S.A. (2017). Sphingolipids and phospholipids in insulin resistance and related metabolic disorders. Nat. Rev. Endocrinol. *13*, 79–91. https://doi.org/10.1038/nrendo.2016.169.

34. Palau-Rodriguez, M., Garcia-Aloy, M., Miñarro, A., Bernal-Lopez, M.R., Brunius, C., Gómez-Huelgas, R., Landberg, R., Tinahones, F.J., and Andres-Lacueva, C. (2020). Effects of a long-term lifestyle intervention on metabolically healthy women with obesity: Metabolite profiles according to weight loss response. Clin. Nutr. *39*, 215–224. https://doi.org/10.1016/j.clnu.2019.01.018.

35. Shah, R.V., Steffen, L.M., Nayor, M., Reis, J.P., Jacobs, D.R., Allen, N.B., Lloyd-Jones, D., Meyer, K., Cole, J., Piaggi, P., et al. (2023). Dietary metabolic signatures and cardiometabolic risk. Eur. Heart J. *44*, 557–569. https://doi.org/10.1093/eurheartj/ehac446.

36. Short, K.R., Chadwick, J.Q., Teague, A.M., Tullier, M.A., Wolbert, L., Coleman, C., and Copeland, K.C. (2019). Effect of Obesity and Exercise Training on Plasma Amino Acids and Amino Metabolites in American Indian Adolescents. J. Clin. Endocrinol. Metab. *104*, 3249–3261. https://doi.org/10.1210/jc.2018-02698.

37. Feng, Y., Fu, M., Guan, X., Wang, C., Yuan, F., Bai, Y., Meng, H., Li, G., Wei, W., Li, H., et al. (2021). Uric Acid Mediated the Association Between BMI and Postmenopausal Breast Cancer Incidence: A Bidirectional Mendelian Randomization Analysis and Prospective Cohort Study. Front. Endocrinol. *12*, 742411. https://doi.org/10.3389/fendo.2021.742411.

38. Liang, Y., Dai, X., Cao, Y., Wang, X., Lu, J., Xie, L., Liu, K., and Li, X. (2023). The neuroprotective and antidiabetic effects of trigonelline: A review of signaling pathways and molecular mechanisms. Biochimie *206*, 93–104. https://doi.org/10.1016/j.biochi.2022.10.009.

39. Bjørngaard, J.H., Nordestgaard, A.T., Taylor, A.E., Treur, J.L., Gabrielsen, M.E., Munafò, M.R., Nordestgaard, B.G., Åsvold, B.O., Romundstad, P., and Davey Smith, G. (2017). Heavier smoking increases coffee consumption: findings from a Mendelian randomization analysis. Int. J. Epidemiol. *46*, 1958–1967. https://doi.org/10.1093/ije/dyx147.

40. Treur, J.L., Taylor, A.E., Ware, J.J., McMahon, G., Hottenga, J.-J., Baselmans, B.M.L., Willemsen, G., Boomsma, D.I., Munafò, M.R., and Vink, J.M. (2016). Associations between smoking and caffeine consumption in two European cohorts. Addiction *111*, 1059–1068. https://doi.org/10.1111/add.13298.

41. Ford, E.S., Zhao, G., Tsai, J., and Li, C. (2011). Low-risk lifestyle behaviors and all-cause mortality: findings from the National Health and Nutrition Examination Survey III Mortality Study. Am. J. Publ. Health *101*, 1922–1929. https://doi.org/10.2105/AJPH.2011.300167.

42. van Dam, R.M., Li, T., Spiegelman, D., Franco, O.H., and Hu, F.B. (2008). Combined impact of lifestyle factors on mortality: prospective cohort study in US women. BMJ *337*, a1440. https://doi.org/10.1136/bmj.a1440.

43. Li, Y., Schoufour, J., Wang, D.D., Dhana, K., Pan, A., Liu, X., Song, M., Liu, G., Shin, H.J., Sun, Q., et al. (2020). Healthy lifestyle and life expectancy free of cancer, cardiovascular disease, and type 2 diabetes: prospective cohort study. BMJ *368*, l6669. https://doi.org/10.1136/bmj.l6669.

44. Tamakoshi, A., Tamakoshi, K., Lin, Y., Yagyu, K., and Kikuchi, S.; JACC Study Group (2009). Healthy lifestyle and preventable death: findings from the Japan Collaborative Cohort (JACC) Study. Prev. Med. *48*, 486–492. https://doi.org/10.1016/j.ypmed.2009.02.017.

45. Sun, Q., Yu, D., Fan, J., Yu, C., Guo, Y., Pei, P., Yang, L., Chen, Y., Du, H., Yang, X., et al. (2022). Healthy lifestyle and life expectancy at age 30 years in the Chinese population: an observational study. Lancet Public Health *7*, e994–e1004. https://doi.org/10.1016/S2468-2667(22)00110-4.

46. Li, K., Hüsing, A., and Kaaks, R. (2014). Lifestyle risk factors and residual life expectancy at age 40: a German cohort study. BMC Med. *12*, 59. https://doi.org/10.1186/1741-7015-12-59.

47. O'Doherty, M.G., Cairns, K., O'Neill, V., Lamrock, F., Jørgensen, T., Brenner, H., Schöttker, B., Wilsgaard, T., Siganos, G., Kuulasmaa, K., et al. (2016). Effect of major lifestyle risk factors, independent and jointly, on life expectancy with and without

cardiovascular disease: results from the Consortium on Health and Ageing Network of Cohorts in Europe and the United States (CHANCES). Eur. J. Epidemiol. *31*, 455–468. https://doi.org/10.1007/s10654-015-0112-8.

48. López-Otín, C., Galluzzi, L., Freije, J.M.P., Madeo, F., and Kroemer, G. (2016). Metabolic Control of Longevity. Cell *166*, 802–821. https://doi.org/10.1016/j.cell.2016.07.031.

49. Bao, Y., Bertoia, M.L., Lenart, E.B., Stampfer, M.J., Willett, W.C., Speizer, F.E., and Chavarro, J.E. (2016). Origin, Methods, and Evolution of the Three Nurses' Health Studies. Am. J. Publ. Health *106*, 1573–1581. https://doi.org/10.2105/AJPH.2016.303338.

50. Rimm, E.B., Giovannucci, E.L., Willett, W.C., Colditz, G.A., Ascherio, A., Rosner, B., and Stampfer, M.J. (1991). Prospective study of alcohol consumption and risk of coronary disease in men. Lancet *338*, 464–468. https://doi.org/10.1016/0140-6736(91)90542-W.

51. (1998). Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. Contr. Clin. Trials *19*, 61–109. https://doi.org/10.1016/s0197-2456(97)00078-0.

52. Vaughan, L., Espeland, M.A., Snively, B., Shumaker, S.A., Rapp, S.R., Shupe, J., Robinson, J.G., Sarto, G.E., and Resnick, S.M.; Women's Health Initiative Memory Study of Younger Women (WHIMS-Y) Study Group (2013). The rationale, design, and baseline characteristics of the Women's Health Initiative Memory Study of Younger Women (WHIMS-Y). Brain Res. *1514*, 3–11. https://doi.org/10.1016/j.brainres.2013.03.047.

53. Paynter, N.P., Balasubramanian, R., Giulianini, F., Wang, D.D., Tinker, L.F., Gopal, S., Deik, A.A., Bullock, K., Pierce, K.A., Scott, J., et al. (2018). Metabolic Predictors of Incident Coronary Heart Disease in Women. Circulation *137*, 841–853. https://doi.org/10.1161/CIRCULATIONAHA.117.029468.

54. Chu, N.-F., Spiegelman, D., Yu, J., Rifai, N., Hotamisligil, G.S., and Rimm, E.B. (2001). Plasma leptin concentrations and four-year weight gain among US men. Int. J. Obes. *25*, 346–353. https://doi.org/10.1038/sj.ijo.0801549.

55. Hankinson, S.E., Willett, W.C., Manson, J.E., Colditz, G.A., Hunter, D.J., Spiegelman, D., Barbieri, R.L., and Speizer, F.E. (1998). Plasma Sex Steroid Hormone Levels and Risk of Breast Cancer in Postmenopausal Women. J. Natl. Cancer Inst. *90*, 1292–1299. https://doi.org/10.1093/jnci/90.17.1292.

56. Wang, F., Baden, M.Y., Guasch-Ferré, M., Wittenbecher, C., Li, J., Li, Y., Wan, Y., Bhupathiraju, S.N., Tobias, D.K., Clish, C.B., et al. (2022). Plasma metabolite profiles related to plant-based diets and the risk of type 2 diabetes. Diabetologia *65*, 1119–1132. https://doi.org/10.1007/s00125-022-05692-8.

57. Fiehn, O., Robertson, D., Griffin, J., van der Werf, M., Nikolau, B., Morrison, N., Sumner, L.W., Goodacre, R., Hardy, N.W., Taylor, C., et al. (2007). The metabolomics standards initiative (MSI). Metabolomics *3*, 175–178. https://doi.org/10.1007/s11306-007-0070-6.

58. Kim, H.-Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. Restor. Dent. Endod. *38*, 52–54. https://doi.org/10.5395/rde.2013.38.1.52.

59. Wei, R., Wang, J., Su, M., Jia, E., Chen, S., Chen, T., and Ni, Y. (2018). Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data. Sci. Rep. *8*, 663. https://doi.org/10.1038/s41598-017-19120-0.

60. Willett, W.C., Browne, M.L., Stampfer, M.J., Bain, C., Bain, C., Rosner, B., Colditz, G.A., Hennekens, C.H., Speizer, F.E., Stampfer, M.J., et al. (1985). REPRODUCIBILITY AND VALIDITY OF A SEMIQUANTITATIVE FOOD FREQUENCY QUESTIONNAIRE. Am. J. Epidemiol. *122*, 731–740.

61. Wolf, A.M., Hunter, D.J., Colditz, G.A., Manson, J.E., Stampfer, M.J., Corsano, K.A., Rosner, B., Kriska, A., and Willett, W.C. (1994). Reproducibility and validity of a self-administered physical activity questionnaire. Int. J. Epidemiol. *23*, 991–999. https://doi.org/10.1093/ije/23.5.991.

62. Balasubramanian, R., Paynter, N.P., Giulianini, F., Manson, J.E., Zhao, Y., Chen, J.-C., Vitolins, M.Z., Albert, C.A., Clish, C., and Rexrode, K.M. (2020). Metabolomic profiles associated with all-cause mortality in the Women's Health Initiative. Int. J. Epidemiol. *49*, 289–300. https://doi.org/10.1093/ije/dyz211.

63. Lee, L.O., James, P., Zevon, E.S., Kim, E.S., Trudel-Fitzgerald, C., Spiro, A., Grodstein, F., and Kubzansky, L.D. (2019). Optimism is associated with exceptional longevity in 2 epidemiologic cohorts of men and women. Proc. Natl. Acad. Sci. USA *116*, 18357–18362. https://doi.org/10.1073/pnas.1900712116.

64. Wang, F., Tessier, A.-J., Liang, L., Wittenbecher, C., Haslam, D.E., Fernández-Duval, G., Heather Eliassen, A., Rexrode, K.M., Tobias, D.K., Li, J., et al. (2023). Plasma metabolomic profiles associated with mortality and longevity in a prospective analysis of 13,512 individuals. Nat. Commun. *14*, 5744. https://doi.org/10.1038/s41467-023-41515-z.

65. Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., and Sergushichev, A. (2021). Fast gene set enrichment analysis. Preprint at bioRxiv. https://doi.org/10.1101/060012.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Blood plasma	This study	N/A
Software and algorithms		
R and R Studio	R Project for Statistical Computing	https://www.r-project.org/
SAS software	SAS Institute	https://www.sas.com/en_us/home.html
Analytical code	Zenodo GitHub	https://doi.org/10.5281/zenodo.10359825

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Marta Guasch-Ferré (mguasch@hsph.harvard.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The participants data reported in this study cannot be deposited in a public repository due to participant confidentiality and privacy concerns. Therefore, data are available upon written request. According to standard controlled access procedure, applications to use the NHSI, NHSII and HPFS resources will be reviewed by our External Collaborators Committee for scientific aims, evaluation of the fit of the data for the proposed methodology, and verification that the proposed use meets the guidelines of the Ethics and Governance Framework and of the consent that was provided by the participants. Further information including the procedures for obtaining and accessing data from the Nurses' Health Studies and Health Professionals' Follow-up Study is described at https://www.nurseshealthstudy.org/researchers and https://sites.sph.harvard.edu/hpfs/for-collaborators. To request access, contact nhsaccess@channing.harvard.edu. All original analytical code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Primary analyses were performed in the NHS, NHSII, and HPFS prospective cohort studies. The NHS began in 1976 and recruited 121,700 U.S. women registered nurses aged 30–55 years[49]; the NHSII was initiated in 1989 and enrolled 116,429 women registered nurses 25–42 years of age[49]; and the HPFS, established in 1986, recruited 51,525 men health professionals aged 40–75 years.[50] The cohorts have been described in detail.[49,50] Blood samples were collected from 32,826 NHS participants between 1989 and 1990, 29,611 NHSII participants between 1996 and 1999, and 18,225 HPFS participants between 1993 and 1995. From these population samples, 14 nested case-control studies were previously conducted for metabolomic profiling.

For the current study, the baseline was set as the respective blood draw date for each participant. Participants with a history of CVD or cancer at baseline, who reported

implausible energy intakes (men: <800 or >4,200 kcal/d; women: <600 or >3,500 kcal/d), or with unavailable metabolomics profiling, dietary intakes, alcohol intake, physical activity, BMI, or smoking were excluded. A total of 11,487 participants remained in the cross-sectional metabolomic analyses and 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 replication study was performed in a nested case-control study of coronary heart disease within the WHI. The WHI study was initiated in 1993 and enrolled 161,808 U.S. postmenopausal women aged 50 to 79 years[51,52] in an observational study (WHI-OS) or randomized controlled trials of hormonal therapy (WHI-HT). Participants completed baseline socio-demographic, diet, lifestyle, and medical history questionnaires. Specifically, the nested case-control study included 2,306 participants with blood collected at enrollment.[53] Individuals with a history of CVD, cancer (excluding non-melanoma skin cancer), without metabolomic profiling, and with missing dietary intake data or lifestyle questionnaire at baseline (blood draw) were excluded (n = 737). A total of 1,569 participants were included in the replication analyses. The protocol was approved by the Fred Hutchinson Cancer Research Center Institutional Review Board, Seattle, WA. Informed consent was obtained from all participants. The study flow diagram is represented in Figure 1.

METHOD DETAILS

Blood collection and metabolite profiling

Participants were sent a blood collection kit and arranged to have their blood sample drawn. Blood was collected in sodium heparin (NHS and NHSII) or EDTA (HPFS) vacutainers, mailed in an enclosed ice pack via overnight courier and centrifuged upon arrival at the laboratory.[54,55] Plasma was aliquoted and kept in continuously monitored liquid nitrogen freezers ($\leq -130^{\circ}C$) until analysis. Seventy-two percent of NHSI/NHSII/HPFS participants provided fasting (≥ 8 h) blood samples.

Plasma metabolomic profiling was performed at the Broad Institute of the MIT and Harvard (Cambridge, MA, USA) using a liquid chromatography-mass spectrometry (LC-MS) as described elsewhere.[53,56] For amino acids and other polar compounds, metabolites were extracted using acetonitrile/methanol/formic acid with the internal standards valine-d8 and phenylalanine-d8; then, separated through hydrophilic interaction liquid chromatography (HILIC; Atlantis HILIC column, Waters, MA, USA) and analyzed by positive ionization mode MS (Thermo Fisher Scientific, MA, USA). For lipids, isopropanol was used for extraction with the internal standard 1,2- didodecanoyl-*sn*-glycero-3-phosphocholine. Fractions were separated by octyl high-performance liquid chromatography (HPLC; ACQUITY BEH C8 column; Waters, MA, USA), and metabolites were analyzed by positive ionization mode MS. Targeted raw metabolites were processed using TraceFinder 3.3 software (Thermo Fisher Scientific; Waltham, MA), and non-targeted raw metabolites using Progenesis QI (Nonlinear Dynamics; Newcastle upon Tyne, UK). Metabolites were identified as per the Metabolite Standard Initiative (MSI).[57] In the WHI, plasma metabolites were profiled using the same method.[53]

In total, 404 metabolites of known identity were measured. Metabolites for which quality control replicates intraclass correlation coefficient was <0.3 (n = 8) and detection rate was <75% (n = 153) were excluded. The final number of metabolites considered in the primary analysis was 243. Metabolites predominantly consisted of lipids

including glycerolipids (28.8%), glycerophospholipids (12.3%), plasmalogens (10.3%), carnitines (8.2%), cholesterol esters (CE; 4.1%), lysophospholipids (3.3%) and sphingolipids (1.6%); but also, amines and amino acids (12.3%), nucleosides and derivatives (4.9%), and other metabolites (14.0%). In the WHI study, 213 of these metabolites were available for replication analysis. For all cohorts, metabolites with skewness above or below 2 were log-transformed.[58] Within each laboratory analysis batch, metabolites were standardized to z-scores with a mean of 0 and standard deviation of 1. Missing metabolites were imputed using the random forest imputation method.[59]

Assessment of lifestyle factors

Adherence to a healthy lifestyle was evaluated based on five self-reported lifestyle factors: diet, alcohol consumption, physical activity, smoking, and BMI. Diet and alcohol were assessed using an extensively validated 149-item food frequency questionnaire.[60] Diet quality was evaluated using the Alternative Healthy Eating Index (AHEI), excluding alcohol. A healthier diet was defined as an AHEI score $\geq 60^{th}$ percentile of each cohort distribution.[3,43] Moderate alcohol consumption was defined as 5–15 g/d for women and 5–30 g/d for men. Physical activity levels were assessed using a validated questionnaire.[61] A healthy physical activity level was defined as >30 min/d (3.5 h/week) of moderate or vigorous activities requiring ≥ 3 metabolic equivalents of task per hour, including brisk walking ("How much physical activity do adults need?", Centers for Disease Control and Prevention). BMI was calculated as self-reported weight in kilograms, divided by height in meters; and a BMI within the "normal" range of 18.5–24.9 kg/m^2 ("Obesity: preventing and managing the global epidemic", report of a WHO consultation) was considered as healthy. The average of the two repeated measurements closest to the blood draw was used for diet, alcohol, physical activity, and BMI. The last measurement recorded was carried forward for non-respondents to both questionnaires. Lastly, never smoking, assessed closest to blood draw, was considered as healthy.

Participants were attributed a score of 1 for each healthy factor that they met and were otherwise attributed a score of 0. The sum of the five factors was computed and ranged from 0 to 5, a higher score indicating a higher adherence to a healthier lifestyle. This widely recognized healthy lifestyle score has been strongly associated with mortality and life expectancy in previous publications.[3,43]

Ascertainment of deaths and longevity

Deaths were mostly reported by a family member or through postal authorities in response to the follow-up questionnaire. The State vital statistics records and the National Death Index were also searched to identify deaths, with >98% of follow-up completion reached in the cohorts. The death cause was confirmed by study physicians' review of medical records, autopsy reports, or death certificates. The International Classification of Diseases, Eighth Revision (ICD-8) in NHS/NHSII and ICD-9 in HPFS, which were the ICD systems used at the time the cohorts began. Longevity was defined as living to ≥ 85 years. This threshold was selected based on the age distribution of our population and previous studies[62–64] also report the use of this definition.

Assessment of covariates

In NHS/NHSII/HPFS, participants were asked to self-report on race/ethnicity, medication and multivitamin use, diabetes, hypertension, and hypercholesterolemia during follow-up preceding and following the blood draw. The value closest to the blood draw was used in analyses. In WHI, socio-demographic and medical history

questionnaires were collected at the same time as the blood draw, during the initial visit.

QUANTIFICATION AND STATISTICAL ANALYSIS

In NHS/NHSII/HPFS, the associations of individual metabolites (per 1-SD increment) with the combined self-reported healthy lifestyle score were examined using linear regressions, and associations with individual lifestyle factors (adhering to the healthy levels of each factor; yes/no) using logistic regressions. Models were adjusted for baseline age, fasting status, ethnicity, multivitamin use, diabetes, hypertension, antihypertensive medication use, hypercholesterolemia, lipid-lowering medication use, total energy intake, study cohorts, original sub-studies, and the case/control status within the original sub-study. Models of the associations with individual healthy lifestyle factors were further adjusted for the other factors. Subgroup analyses were also conducted by baseline age ($<$ or $\geq$55 years), sex (women or men), and case-control status. Results were corrected for multiple comparison using the Bonferroni adjustment (p value/243 metabolites).

Based on chemical similarity, individual metabolites were grouped into 19 predefined sets of metabolites (https://doi.org/10.5281/zenodo.10359825, MSEA). MSEA was used to identify the metabolite sets most strongly positively and inversely associated with the self-reported healthy lifestyle score. This analysis uses the β coefficients obtained from the multivariable-adjusted linear regression of the association between individual metabolites and the self-reported healthy lifestyle score to estimate p-values for each metabolite set using an adaptive multi-level split Monte-Carlo scheme.[65] p-values were corrected for multiple comparison using the Benjamini-Hochberg FDR adjustment.

The healthy lifestyle metabolomic signature was conducted using elastic net penalized linear regressions. Briefly, this machine learning method regularizes the statistical model by accounting for both the Lasso and Ridge penalties, thereby controlling for variance and bias. Training and test sets with a 70-30 split were used to avoid overfitting of the model; participants were randomized to either group. First, the elastic net model (alpha = 0.5) was developed in the training set using a 10-fold cross-validation technique. The lambda that minimized the cross-validation prediction error rate was selected (lambda = 0.021). Second, the metabolomic signature was computed in the test set by taking the weighted sum of selected metabolites, with the training set-derived elastic net regression coefficients as the corresponding weights. Lastly, in the training set, the signature was obtained using a leave-one-out approach. Correlations between the healthy lifestyle metabolomic signature and the self-reported healthy lifestyle score were examined using the Pearson correlation coefficient (Point-biserial correlation coefficient for individual lifestyle factors) overall, by training-test sets and by cohorts.

In NHS/NHSII/HPFS, the prospective associations of the healthy lifestyle metabolomic signature with all-cause mortality, cardiovascular and cancer mortality, and with longevity were assessed using Cox proportional hazard ratio models. In the longevity analysis, the timescale was censored to surviving to the age of 85 years, death was used as the event, and participants who were alive and younger than 85 years at the end of the follow-up were right censored. To interpret the results per one standard deviation unit increment and allow comparison, we standardized the signature and the self-reported healthy lifestyle scores to Z-scores. Model 1 was adjusted for fasting status (yes or no), and stratified by baseline age, study

cohorts, original sub-studies, and case-control status within the original sub-study. Model 2 was further adjusted for race/ethnicity (white or non-white), multivitamin use (yes or no), diabetes (yes or no), hypercholesterolemia (yes or no), lipid-lowering medication use (yes or no), hypertension (yes or no), antihypertensive medication use (yes or no) and total energy intake (continuous). Model 3 was additionally adjusted for the self-reported healthy lifestyle score (continuous) to examine the independent association of its metabolomic signature. For each participant, the follow-up period was calculated from baseline until the last questionnaire returned (end of follow-up: June 2018 for the NHS/HPFS and June 2019 for the NHSII) or date of death, whichever came first. Based on the results of the Schoenfeld residuals test, there was no evidence to suggest a violation of the proportional hazards assumption in any of the models examined. Additionally, mediation analyses were conducted to estimate the percentage of the associations between the self-reported healthy lifestyle score and mortality and longevity, that can be attributed to the healthy lifestyle metabolomic profile. Mediation proportions and their 95%CI were calculated using the publicly available % Mediate SAS macro (https://ysph.yale.edu/cmips/research/software/mediate_340185_284_47911_v2.pdf).

We performed several sensitivity analyses. We ran the main multivariable-adjusted linear regression models of the associations between individual metabolites and the healthy lifestyle score, without adjusting for potential biological intermediates including diabetes, hypertension, hypercholesterolemia, and related medication use. Also, given the mixed evidence of alcohol consumption on health outcomes, we ran the main models using the self-reported healthy lifestyle score without the alcohol component (score ranging from 0 to 4). Furthermore, we replicated the main analyses in the WHI to validate our findings in an external and independent cohort.

Analyses were performed using R version 4.1.0. The R package "missRanger" was used for imputation of metabolites, "glmnet" for elastic net regression, "survival" for Cox proportional-hazards models and "fgsea" for MSEA. Statistical tests were two-sided and p values <0.05 were considered for significance.