Metabolomic Profiles Associated With Incident Ischemic Stroke

Raji Balasubramanian, ScD, Jie Hu, MD, PhD, Marta Guasch-Ferre, PhD, Jun Li, MD, PhD, Farzaneh Sorond, MD,
Yibai Zhao, MS, Katherine H. Shutta, MS, Jordi Salas-Salvado, MD, PhD, Frank Hu, MD, PhD, Clary B. Clish, PhD,
and Kathryn M. Rexrode, MD, MPH

Correspondence
Dr. Balasubramanian
rbalasub@umass.edu

MORE ONLINE

Class of Evidence
Criteria for rating
therapeutic and diagnostic
studies
NPub.org/coe

Abstract

Background and Objectives

Women have higher lifetime risk of stroke than men, and metabolic factors seem more strongly associated with stroke for women than men. However, few studies in either men or women have evaluated metabolomic profiles and incident stroke.

Methods

We applied liquid chromatography–tandem mass spectrometry to measure 519 plasma metabolites in a discovery set of women in the Nurses' Health Study (NHS; 454 incident ischemic stroke cases, 454 controls) with validation in 2 independent, prospective cohorts: Prevención con Dieta Mediterránea (PREDIMED; 118 stroke cases, 791 controls) and Nurses' Health Study 2 (NHS2; 49 ischemic stroke cases, 49 controls). We applied logistic regression models with stroke as the outcome to adjust for multiple risk factors; the false discovery rate was controlled through the q value method.

Results

Twenty-three metabolites were significantly associated with incident stroke in NHS after adjustment for traditional risk factors ($q < 0.05$). Of these, 14 metabolites were available in PREDIMED and 3 were significantly associated with incident stroke: methionine sulfoxide, N6-acetyllysine, and sucrose ($q < 0.05$). In NHS2, one of the 23 metabolites (glucuronate) was significantly associated with incident stroke ($q < 0.05$). For all 4 metabolites, higher levels were associated with increased risk. These 4 metabolites were used to create a stroke metabolite score (SMS) in the NHS and tested in PREDIMED. Per unit of standard deviation of SMS, the odds ratio for incident stroke was 4.12 (95% confidence interval [CI] 2.26–7.51) in PREDIMED, after adjustment for risk factors. In PREDIMED, the area under the receiver operating characteristic curve (AUC) for the model including SMS and traditional risk factors was 0.70 (95% CI 0.75–0.79) vs the AUC for the model including the traditional risk factors only of 0.65 (95% CI 0.70–0.75), corresponding to a 5% improvement in risk prediction with SMS ($p < 0.005$).

Discussion

Metabolites associated with stroke included 2 amino acids, a carboxylic acid, and sucrose. A composite SMS including these metabolites was associated with ischemic stroke and showed improvement in risk prediction beyond traditional risk factors.

Classification of Evidence

This study provides Class II evidence that a SMS accurately predicts incident ischemic stroke risk.

From the Department of Biostatistics and Epidemiology (R.B., Y.Z., K.H.S.), University of Massachusetts–Amherst; Division of Women's Health (J.H., K.M.R.) and Channing Division of Network Medicine, Department of Medicine (M.G.-F., F.H.), Brigham and Women's Hospital, Harvard Medical School; Departments of Nutrition (M.G.-F., J.L., F.H.) and Epidemiology (J.L., F.H.), Harvard T.H. Chan School of Public Health, Boston, MA; Davee Department of Neurology, Division of Stroke and Neurocritical Care (F.S.), Northwestern Feinberg School of Medicine, Chicago, IL; Departament de Bioquímica i Biotecnologia, Unitat de Nutrició (J.S.S.), Universitat Rovira i Virgili, Reus; Centro de Investigación Biomédica en Red Fisiopatología de la Obesidad y la Nutrición (CIBEROBN) (J.S.-S.), Institute of Health Carlos III, Madrid; Nutrition Unit, Pere Virgili Research Institute (IISPV) (J.S.-S.), University Hospital of Sant Joan de Reus, Spain; and Broad Institute of the Massachusetts Institute of Technology and Harvard University (C.B.C.), Cambridge.

Go to Neurology.org/N for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

Glossary

ARIC = Atherosclerosis Risk in Communities; **AUC** = area under the receiver operating characteristic curve; **BMI** = body mass index; **CHD** = coronary heart disease; **CI** = confidence interval; **CV** = coefficient of variation; **CVD** = cardiovascular disease; **DAG** = diacylglycerol; **FDR** = false discovery rate; **HDL** = high-density lipoprotein; **HT** = hormone therapy; **LC-MS** = liquid chromatography–mass spectrometry; **MetSO** = methionine sulfoxide; **MI** = myocardial infarction; **MSEA** = metabolite set enrichment analysis; **NES** = normalized enrichment score; **NHS** = Nurses' Health Study; **NHS2** = Nurses' Health Study 2; **NMR** = nuclear magnetic resonance; **OR** = odds ratio; **PREDIMED** = Prevención con Dieta Mediterránea; **SMS** = stroke metabolite score; **TAG** = triacylglycerol.

Stroke is the leading cause of long-term disability and the fourth leading cause of death for women in the United States.[1] Women account for almost 60% of all stroke deaths and the lifetime stroke risk is 20% higher in women than men.[1,2] Clear sex differences in stroke exist. Women with stroke have higher rates of hypertension and atrial fibrillation than men and lower rates of prior coronary heart disease (CHD).[1] Diabetes is a stronger risk factor for stroke in women than men.[3] Although we have substantial knowledge of key risk factors for stroke in women, there is still a gap between identified risk factors and the most proximate biological processes that mediate this risk. Improved understanding of the metabolic milieu through metabolomic measurements may advance biologic therapies and identify new pharmacologic strategies for stroke prevention and treatment.

Previous literature has highlighted the associations of metabolites with stroke risk factors including type 2 diabetes[4] and cardiovascular disease (CVD).[5] Previous work has also suggested that n-3 long chain fatty acids may be related to lower risk of stroke.[6,7] We, and others, have previously identified metabolites associated with incident CHD,[8-16] but data for stroke are limited. Analyzing 114 strokes that occurred in 3,904 men and women from the Atherosclerosis Risk in Communities (ARIC) study, investigators identified serum levels of 2 long-chain dicarboxylic acids, tetradecanedioate and hexadecanedioate, as being strongly associated with incident ischemic stroke after adjusting for risk factors.[17] A recent meta-analysis of data from 7 prospective cohorts investigated 147 circulating plasma or serum metabolites using nuclear magnetic resonance and 1,791 incident strokes in 38,797 participants. The analyses revealed 10 significant associations including amino acids (histidine, phenylalanine), glycolysis-related metabolites, acute phase reaction markers, and lipoprotein subfractions.[18]

Despite strong evidence for sex differences in stroke and for sex-specific differences in the serum metabolome,[19] prospective studies of stroke in women are scarce. In order to identify metabolites associated with incident stroke, we examined the association between metabolomic profiles and incident ischemic stroke in female participants of the Nurses' Health Study (NHS). We then validated these stroke metabolites in 2 independent validation datasets: the Nurses' Health Study 2 (NHS2) and the Prevención con Dieta Mediterránea (PREDIMED) prospective[20,21] cohorts. We created a stroke metabolite score (SMS) for incident stroke of the validated metabolites in the NHS and evaluated its associated risk prediction over traditional risk factors in the PREDIMED cohort.

The primary research question being addressed in this study is to discover and validate metabolomic profiles that are associated with incident ischemic stroke risk after adjustment for traditional stroke risk factors utilizing multiple, independent prospective cohorts.

Methods

Datasets

Nurses' Health Study (Discovery Cohort)

The discovery dataset was sampled from the NHS, a prospective cohort study established in 1976, when 121,700 US female registered nurses aged 30–55 years completed a baseline questionnaire. All NHS participants completed biennial questionnaires since study inception to report exposure status and disease diagnoses. From 1989 to 1990, 32,826 NHS participants, ages 43–70 years, provided a heparin blood sample and completed a short questionnaire.[22] From 2000 to 2002, a second blood sample was collected from a subset of these women (n = 18,743 women, aged 53–80 years, >98% postmenopausal) using the same protocol as in the original collection.[23]

From this subcohort with at least one blood specimen, 454 women with incident ischemic stroke confirmed by medical record review (cases) and an equal number of 1:1 matched controls were selected into this study. Matching factors included age, race, hormone therapy (HT) use, smoking status, history of CHD, and fasting status.

Nurses' Health Study 2 (Validation Cohort)

The NHS2 was established in 1989 including a total of 116,430 women, ages 25–42 years at entry into the study. As in the NHS, all NHS2 participants completed biennial questionnaires since study inception to report exposure status and disease diagnoses. Blood samples were collected between 1996 and 1999 in a subset of 29,611 participants ages 32–54 years at blood draw.[24] From this subset of participants, 49

women with confirmed incident ischemic stroke and 49 controls matched by the same methods as NHS were selected.

Our study was restricted to ischemic strokes confirmed by physician review of medical records. Participants who reported a nonfatal stroke on biennial questionnaire were asked for permission to review their medical records. Deaths were reported by next of kin or postal authorities or determined by systematic searches of the National Death Index, and permission for medical records was sought. Strokes were confirmed according to the National Survey of Stroke criteria,[25] requiring evidence of a sudden or rapid onset of neurologic deficit that persisted for more than 24 hours or until death. Strokes were classified as ischemic stroke when there was evidence of thrombotic or embolic occlusion of a cerebral artery. The reproducibility of our classification system is high.[26] Thrombotic strokes were further defined as large artery infarctions (involving the cortical artery regions) or small artery infarctions (in deep areas without involvement of the cortex). Strokes were defined as embolic if evidence of an embolic source was present in the medical record and if imaging studies or neurology consult supported the diagnosis; large artery infarctions due to carotid atherosclerotic plaque were classified as embolic strokes. CT or MRI documentation was available for 95% of cases.

PREDIMED (Validation Cohort)

The PREDIMED trial[27] randomized 7,447 men and women at high risk of developing CVD to 1 of 3 interventions: a Mediterranean dietary intervention supplemented with virgin olive oil, a Mediterranean dietary intervention supplemented with nuts, or advice about a low-fat diet. All participants provided written informed consent. The primary endpoint for this trial was a composite of myocardial infarction (MI), stroke, and CVD death. The metabolomics substudy utilized a case–cohort study design nested within PREDIMED: a random selection of eligible participants with baseline EDTA plasma samples (approximately 10% of the cohort) and all incident CVD cases with available samples were selected. This PREDIMED dataset of 980 participants included 791 noncases and 118 incident strokes, 76 incident MI, and 5 participants with both incident stroke and MI.[28] We excluded all participants with incident MI and based the analysis on the 791 noncases and 118 incident stroke cases. Plasma samples were collected using EDTA, processed immediately, shipped at −4°C, and then stored at −70°C until analysis.

Metabolomics

For all 3 cohorts, metabolomic measurements were performed at the Broad Institute using 4 complementary liquid chromatography–mass spectrometry (LC-MS) methods (HILIC-positive, HILIC-negative, C8-positive, C18-negative) as described in a previous publication[16] and in the supplement (eAppendix 1, links.lww.com/WNL/B679). For NHS, metabolomic measures included all 4 platforms with 519 measured metabolites and a median coefficient of variation (CV) of 12.9%. In NHS2, measurements included 4 platforms (HILIC-positive, HILIC-negative, C8-positive, C18-negative)[16]; the subset of metabolites that met the threshold for discovery in NHS were analyzed in each of these datasets. For PREDIMED samples, only the HILIC-positive, C8-positive, and HILIC-negative methods were available.[28] For each method, pooled plasma reference samples were included every 20 samples and results were standardized using the ratio of the value of the sample to the value of the nearest pooled reference multiplied by the median of all reference values for the metabolite. For each method, metabolite identities were confirmed using authentic reference standards or reference samples. Blinded split samples were also included and used to calculate the CV for each metabolite.

Statistical Analysis

Metabolite Preprocessing

All metabolites were natural logarithm transformed and standardized to render their distributions with zero mean and unit variance. Missing values below the limit of detection were imputed to equal one-half the lowest observed value. Metabolites with greater than 50% missing values were excluded.

Two-Stage Study Design

Discovery and validation of metabolite–stroke associations was carried out in a 2-stage framework (Figure 1). Metabolites associated with incident stroke were discovered in the NHS dataset (n = 908; 454 ischemic stroke cases). Metabolites that met the threshold for statistical significance in the discovery stage ($q < 0.05$) were carried forward for validation in the NHS2 dataset (n = 98; 49 incident ischemic strokes) and in the PREDIMED dataset (n = 909; 118 incident strokes).
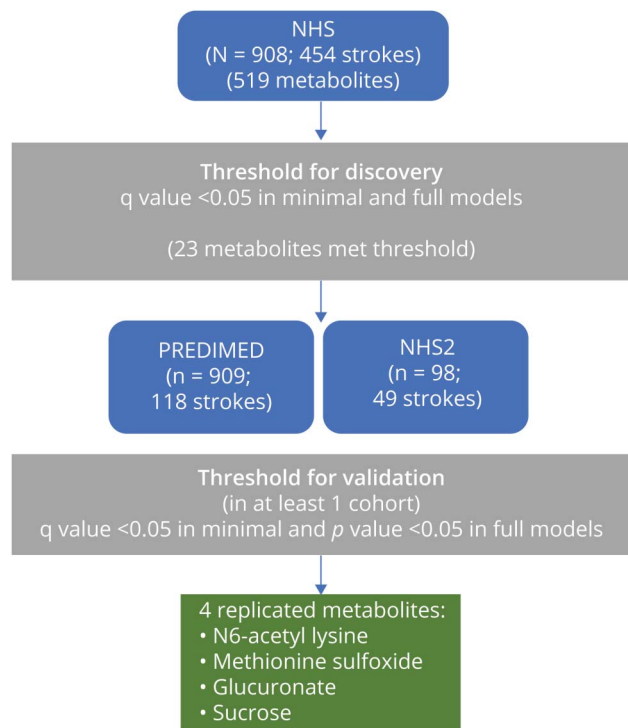
Discovery

In the NHS dataset, 519 identified metabolites were examined individually in conditional logistic models with incident stroke case/control status as outcome and metabolite levels as the primary predictor. Minimal models adjusted only for matching factors including age, race, HT use, smoking status, history of coronary heart disease, and fasting status; fully adjusted models included additional adjustment for body mass index (BMI) (continuous), HT and aspirin use, histories of hypertension, diabetes, and elevated cholesterol, total cholesterol, high-density lipoprotein (HDL), and hemoglobin A1C. Adjustment for multiple testing and control of the false discovery rate (FDR) was based on the Storey q value procedure implemented in the R package *qvalue*.[29] Metabolites that met the q value threshold of 0.05 in the minimal model were evaluated in fully adjusted models. Statistically significant associations in the discovery cohort satisfied a q value <0.05 in both the minimal and fully adjusted models.

Validation

Metabolites that met the threshold for statistical significance in discovery were evaluated individually for association with incident stroke in the NHS2 and PREDIMED datasets. In the

Figure 1 Overview of the 2-Stage Discovery–Validation Framework in 3 Independent Datasets: The Nurses' Health Study (NHS), Nurses' Health Study 2 (NHS2), and Prevención con Dieta Mediterránea (PREDIMED) Studies

Four metabolites (N6-acetyl lysine, methionine sulfoxide, glucuronate, and sucrose) met the threshold for validation in either NHS2 or PREDIMED.

NHS2 dataset, all discovery metabolites were examined individually in conditional logistic regression models using the same covariate sets used for discovery excluding histories of diabetes and elevated cholesterol (due to model convergence failure). In the PREDIMED dataset, the subset of available discovery metabolites was examined individually in logistic regression models with adjustment for age (continuous), sex, intervention arm, BMI (continuous), histories of elevated cholesterol, hypertension, and diabetes, and total and HDL cholesterol. Metabolite interactions with sex were evaluated using the Wald test for the interaction coefficient.

Metabolite associations were considered to have validated if the q value in the minimally adjusted model was less than 0.05 and the p value in the fully adjusted model was less than 0.05 in at least one validation cohort (NHS2, PREDIMED).

Stroke Metabolite Score

The metabolite coefficients in the SMS were estimated in a conditional logistic regression model in the NHS. The conditional logistic regression model included all metabolites that met the threshold for validation in either NHS2 or PREDIMED (Mx_1, Mx_2, Mx_3, Mx_4). The metabolites were natural logarithm transformed and standardized as described earlier.

This set of metabolites was included jointly as predictors with adjustment for the full set of risk factors. The metabolite regression coefficients ($\beta_1, \beta_2, \beta_3, \beta_4$) estimated from the conditional logistic regression model in NHS were used to calculate the SMS for each participant; that is, $SMS = \beta_1*Mx_1 + \beta_2*Mx_2 + \beta_3*Mx_3 + \beta_4*Mx_4$. The SMS was calculated as a measure of the composite effect of the set of validated metabolites. Using the regression coefficients ($\beta_1, \beta_2, \beta_3, \beta_4$) estimated in the NHS cohort, the SMS was calculated similarly for each participant in the NHS2 and PREDIMED. The improvement in risk prediction attributable to the SMS above traditional risk factors was assessed in the PREDIMED cohort. The area under the receiver operating characteristic curve (AUC) associated with the logistic regression model that included SMS and traditional risk factors was compared to the AUC of the model that included traditional risk factors only. The test of the difference in the AUC between these 2 models was based on the DeLong test.[30]

Metabolite set enrichment analysis (MSEA) was performed in the NHS to identify metabolite classes showing coordinated, significant changes in relative abundance between stroke cases and controls. Metabolites were grouped into metabolite sets by mapping 519 metabolites to 26 unique metabolite classes. For each of the 26 metabolite classes, a MSEA enrichment score was calculated to reflect the degree to which the set of metabolites within the class had concordant associations with stroke risk. Details are available in the Supplement (links.lww.com/WNL/B679).

A differential network analysis using the DINGO algorithm[31] was conducted in the NHS to identify the network of differential partial correlations between metabolite pairs, comparing stroke cases to controls. Differential network analysis was carried on the subset of the 519 metabolites measured in the NHS that satisfied a threshold of nominal $p < 0.1$ in fully adjusted models. The DINGO algorithm estimates a metabolomic network in which nodes represent individual metabolites and weighted edges indicate the difference in metabolite pair partial correlations between stroke cases and controls. Thus, a positive edge weight implies that the partial correlation in the stroke cases is higher than in the controls; conversely, a negative edge weight indicates a lower partial correlation in cases relative to controls. Community detection was performed on the differential network to identify densely connected subsets of metabolites within the differential network. The metabolites of the differential network were ranked according to 3 network connectivity measures: hub, betweenness, and closeness centrality, each of which was calculated using the igraph R package.[32] Details are presented in the Supplement, links.lww.com/WNL/B679.

Standard Protocol Approvals, Registrations, and Patient Consents

Approval to conduct this study was received from the Institutional Review Board at Mass General Brigham and all participants gave informed consent.

Data Availability

Qualified researchers may request access to the data from the NHS and NHS2 cohorts by application through the NHS website.[33]

Results

The baseline characteristics of the 908 women in the NHS discovery set, including 454 incident ischemic stroke cases, 98 women in the NHS2 validation set including 49 incident ischemic stroke cases, and 909 participants in the PREDIMED study including 118 incident total strokes are shown in Table 1. The mean age at sample acquisition in the NHS and PREDIMED was comparable (65–70 years); participants in the NHS2 study were younger, with a mean age of 45 years. Median time to incident stroke was 5.5 years in the NHS, 5.2 years in NHS2, and 3.2 years in PREDIMED. In the NHS discovery dataset, a total of 519 unique, identified metabolites were measured; the median (10th–90th percentile) of the CVs was 12.9 (8.8–33.6).

In the NHS, 73 of the 519 metabolites were significantly associated with incident ischemic stroke (q value < 0.05) controlling for matching variables alone (age, race, HT use, smoking status, history of CHD, and fasting status); of these, 23 metabolites remained significant after additionally adjusting for stroke risk factors including BMI, history of elevated cholesterol, history of hypertension, history of diabetes, hormone therapy use, aspirin use, total cholesterol, HDL cholesterol, and hemoglobin A1C (q value < 0.05). Results of the discovery analysis are shown in eTables 1 and 2 (links.lww.com/WNL/B679).

All metabolites that met the threshold for statistical significance in discovery were individually evaluated for association with incident stroke risk in the PREDIMED and NHS2 datasets. Of the 23 metabolites, 14 were available to test in PREDIMED: of these, 3 metabolites were significantly associated with stroke in both the minimally and fully adjusted models (Figure 2, eTable 3, links.lww.com/WNL/B679). N6-acetyllysine, methionine sulfoxide, and sucrose/lactose/trehalose were each associated with increased risk of incident stroke in fully adjusted models, with odds ratios (ORs) for a 1 SD increase in metabolite levels of 1.30 (95% confidence interval [CI] 1.07–1.59), 1.53 (95% CI 1.28–1.83), and 1.31 (95% CI 1.08–1.59), respectively. The compound labeled as sucrose/lactose/trehalose primarily reflects dietary sugar intake. In NHS2, glucuronate was significantly associated with ischemic stroke, with a fully adjusted OR of 2.35 (95% CI 1.13–4.89) for a 1 SD increase in metabolite levels. In addition, in NHS2, the statistical significance of sebacate (also known as decanedioate) was close to the threshold

Table 1 Baseline Characteristics of Participants in the NHS (n = 908), NHS2 (n = 98), and PREDIMED (n = 909) Datasets

	NHS (n = 908)		NHS2 (n = 98)		PREDIMED (n = 909)	
	Ischemic stroke cases	Controls	Ischemic stroke cases	Controls	Stroke cases	Noncases
Number	454	454	49	49	118	791
% Women	100	100	100	100	50.0	56.8
Time to stroke, y	5.5 (2.8)	—	5.2 (3.1)	—	3.2 (1.7)	—
Age, y	65.5 (7.7)	65.5 (7.6)	45.4 (4.0)	45.4 (3.9)	69.9 (6.2)	67.2 (6.0)
BMI, kg/m²	26.5 (5.4)	25.6 (4.5)	28.8 (7.0)	26.7 (7.3)	29.6 (3.7)	29.8 (3.6)
Total cholesterol, mg/dL	220.6 (40.6)	219.8 (42.1)	196.3 (38.4)	196.8 (40.0)	206.7 (34.3)	207.5 (35.5)
HDL cholesterol, mg/dL	59.3 (18.8)	62.6 (18.2)	59.0 (13.0)	61.9 (14.3)	49.4 (10.6)	51.1 (11.5)
Hemoglobin A1C, %	5.9 (1.2)	5.6 (0.6)	6.1 (1.5)	5.5 (0.3)	NA	NA
White	97.1	98.5	95.9	100	100	100
Current smoker	13.2	11.9	14.3	18.4	15.2	12.9
Prevalent diabetes	15.6	7.9	16.3	0	60.7	46.9
Prevalent hypertension	45.8	38.8	16.3	4.1	83.9	83.9
Elevated cholesterol	17.0	13.4	16.3	2.0	39.8	73.6
HT use	48.9	49.3	12.2	18.4	NA	NA
Aspirin use	53.5	50.4	20.4	10.2	NA	NA

Abbreviations: BMI = body mass index; HDL = high-density lipoprotein; HT = hormone therapy; NHS = Nurses' Health Study; NHS2 = Nurses' Health Study 2; PREDIMED = Prevención con Dieta Mediterránea.
Values are mean (SD) or %.

Figure 2 Metabolite Associations With Incident Stroke Discovered in the Nurses' Health Study (NHS) and Validated in Either Nurses' Health Study 2 (NHS2) or Prevención con Dieta Mediterránea (PREDIMED)
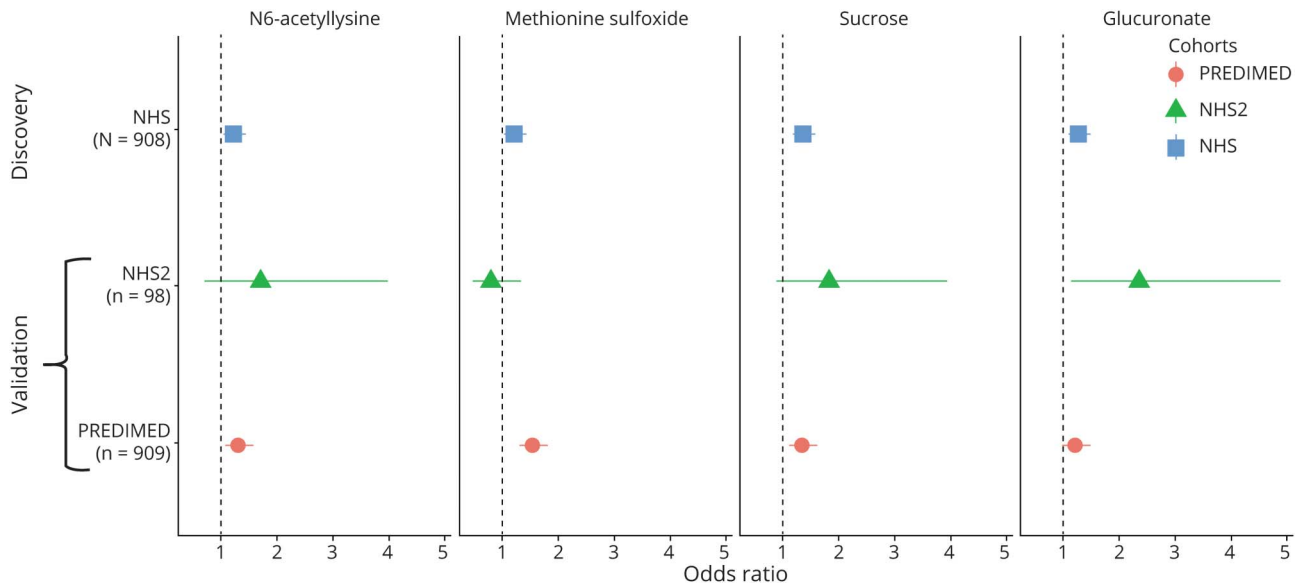
Results from fully adjusted models are shown. NHS/NHS2 models included matching factors (age, race, hormone therapy [HT] use, smoking status, history of coronary heart disease, and fasting status) with additional adjustment for body mass index (BMI), history of elevated cholesterol (NHS only), history of hypertension, history of diabetes (NHS only), hormone therapy use, aspirin use, total cholesterol, high-density lipoprotein (HDL) cholesterol, and hemoglobin A1C. PREDIMED models included age, sex, and intervention arm with additional adjustment for BMI, history of elevated cholesterol, history of hypertension, history of diabetes, total cholesterol, and HDL cholesterol.
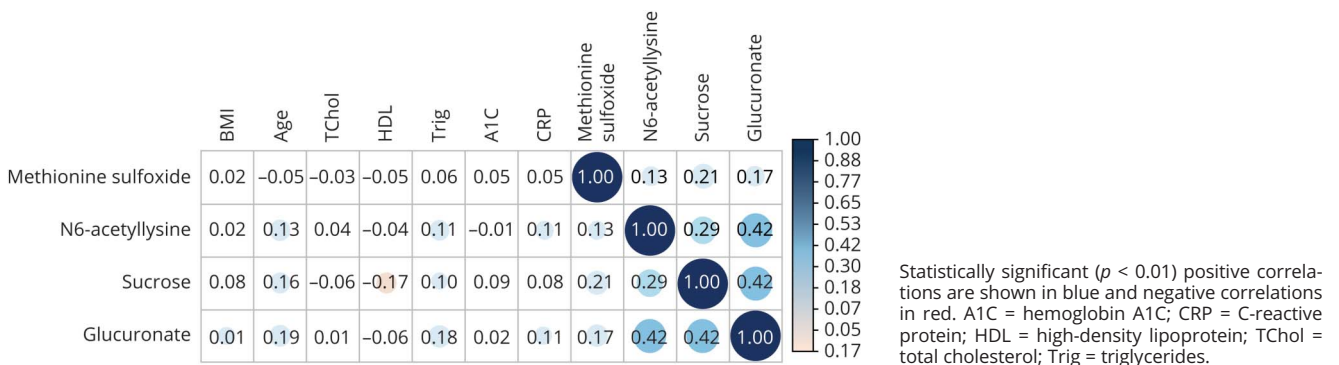
for validation (q value = 0.08) after controlling for matching variables and was associated with risk in the fully adjusted model with an OR of 2.85 (95% CI 1.09–7.47) for a 1 SD increase in levels (eTable 3).

Pearson correlations (r) between the 4 validated metabolites and biomarkers (total and HDL cholesterol, triglycerides, hemoglobin A1C, C-reactive protein) or stroke risk factors (age, BMI) in the NHS are shown in Figure 3. The metabolites were only weakly correlated in magnitude with biomarkers and stroke risk factors (all $r < 0.20$), although several correlations of metabolites with biomarkers were statistically significant ($p < 0.05$). The strongest associations in magnitude were the correlations of glucuronate with age ($r = 0.19$; $p < 10^{-8}$) and with triglyceride levels ($r = 0.18, p < 10^{-6}$). The 4 metabolites were modestly to moderately correlated with one another ($r = 0.13–0.42$). CV and percentage of measurements below the limit of detection for the validated metabolites are provided in eTable 4 (links.lww.com/WNL/B679).

An SMS was estimated in the NHS, in a conditional logistic regression model that simultaneously adjusted for the 4 validated metabolites (N6-acetyllysine, methionine sulfoxide, sucrose/lactose/trehalose, and glucuronate), in addition to

Figure 3 Pearson Correlations Between Validated Metabolites (N6-Acetyllysine, Methionine Sulfoxide, Sucrose, and Glucuronate) and Body Mass Index (BMI), Age, and Continuous Biomarkers in the Nurses' Health Study (NHS)

Statistically significant ($p < 0.01$) positive correlations are shown in blue and negative correlations in red. A1C = hemoglobin A1C; CRP = C-reactive protein; HDL = high-density lipoprotein; TChol = total cholesterol; Trig = triglycerides.

the full set of stroke risk factors. The estimated regression coefficients corresponding to the levels of N6-acetyllysine, methionine sulfoxide, sucrose/lactose/trehalose, and glucuronate in the SMS were 0.09, 0.11, 0.11, and 0.22, respectively. The SMS, treated as a composite measure of the set of validated metabolites, was used to quantify improvements in risk prediction in the PREDIMED cohort. The SMS was associated with incident stroke in PREDIMED with an OR of 4.12 (95% CI 2.26–7.51), corresponding to a 1 SD increase in SMS, after adjusting for traditional risk factors (eTable 5, links.lww.com/WNL/B679). The logistic regression model that included the SMS and traditional risk factors in PREDIMED had an AUC of 0.70 (95% CI 0.75–0.79). In comparison, the AUC for the logistic regression model with the traditional risk factors alone was 0.65 (95% CI 0.70–0.75). This 5% absolute increase in AUC attributable to the SMS above and beyond traditional risk factors was significant ($p < 0.005$).

In the NHS, the 454 incident stroke cases included 125 thrombotic large artery infarcts, 156 thrombotic lacunar infarcts, and 124 embolic strokes. The ORs for the association of SMS with specific stroke subtype were as follows: thrombotic large artery infarct (OR 2.23, 95% CI 0.95–5.23, $p = 0.07$), thrombotic lacunar infarct (OR 5.03, 95% CI 1.84–13.75, $p = 0.002$), and embolic strokes (OR 2.12, 95% CI 0.88–5.12, $p = 0.10$).

The PREDIMED cohort included men (n = 401, 59 stroke cases) and women (n = 508, 59 stroke cases). The SMS was associated with incident stroke risk in analyses stratified by sex, with ORs of 5.42 (95% CI 2.37–12.37) in women and 2.68 (95% CI 1.10–6.53) in men after adjustment for the full set of stroke risk factors. There was no evidence of modification of the association of SMS with incident stroke risk by sex ($p = 0.12$). We also evaluated each of the 4 validated metabolites for effect modification by biological sex in the PREDIMED cohort. There was no evidence of sex differences for N6-acetyllysine ($p = 0.18$), glucuronate ($p = 0.58$), and sucrose ($p = 0.85$). However, there was significant effect modification by sex for methionine sulfoxide in both minimal and fully adjusted models ($p \leq 0.02$). In fully adjusted models, methionine sulfoxide was strongly associated with stroke risk in women (OR 1.85, 95% CI 1.44–2.37, $p < 10^{-5}$); however, this association did not meet the threshold for statistical significance in men (OR 1.14, 95% CI 0.87–1.51, $p = 0.34$).

To evaluate concerted differences in metabolite abundance between stroke cases and controls by metabolite class, MSEA was performed. Three metabolite classes (steroids and derivatives, triacylglycerols [TAGs], and diacylglycerols [DAGs]) were significantly enriched for association with stroke using results from the minimally adjusted model. The metabolite class of steroids and derivatives, with 33 metabolites, had a significant enrichment of metabolites with inverse associations with stroke risk (normalized enrichment score [NES] −2.46, $p < 10^{-5}$, FDR $p < 10^{-4}$). TAGs and DAGs, including 76 and 10 metabolites, respectively, showed

enrichment of metabolites with positive associations with stroke risk (TAG: NES 1.95, $p < 10^{-5}$, FDR $p < 10^{-4}$; DAG: NES = 1.87, $p < 10^{-3}$, FDR $p < 10^{-2}$). However, when results from fully adjusted models were evaluated in the MSEA algorithm, none of the metabolite classes met the FDR $p < 0.05$ threshold for statistical significance. Organooxygen compounds, lysophosphatidylethanolamines, and phosphatidylethanolamine plasmalogens had positive enrichment scores and indoles and derivatives had a negative enrichment score, with a nominal p value threshold of 0.05 (see eFigure 1).

A differential network analysis was carried out to understand the differences in metabolite networks between individuals who developed ischemic stroke vs controls. The DINGO algorithm was used to estimate a differential network comparing the partial correlation networks between stroke cases and controls. Sixty-eight metabolites met the criteria for inclusion in the DINGO algorithm and the resulting differential network included 16 metabolites and 28 differential edges that met the threshold of FDR p value < 0.05. A community detection algorithm applied to the differential network detected 6 disjoint metabolite communities (Figure 4). Results for the 3 centrality measures (hub centrality,[34] betweenness centrality, and closeness centrality[35]) are shown in eTable 6 (links.lww.com/WNL/B679). Three lipid metabolites, including 2 lysophosphatidylethanolamines (C18:3, C16:0) and C44:2 TAG, were central to the differential network and ranked among the top 5 metabolites in the network according to betweenness and closeness centrality.
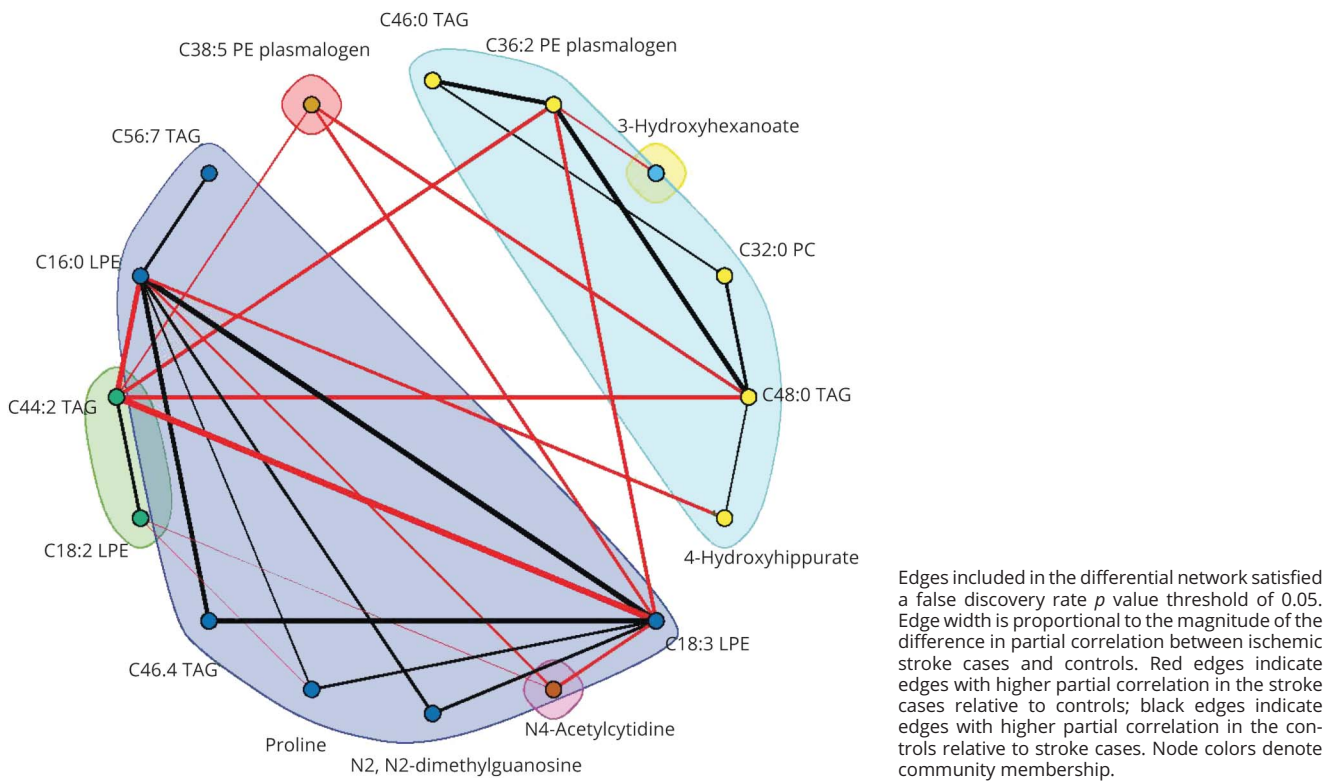
This study provides Class II evidence that a stroke metabolic score accurately predicts incident ischemic stroke risk.

Discussion

In this study examining the association of metabolomic profiles with incident ischemic stroke in women, we discovered and externally validated 4 metabolites including N6-acetyllysine, methionine sulfoxide, sucrose/lactose/trehalose, and glucuronate that remained significantly associated with stroke after adjustment for risk factors. An SMS comprising the validated metabolites was derived in the NHS, and inclusion of the SMS resulted in a significant improvement in stroke risk prediction above traditional stroke risk factors in the PREDIMED cohort.

In our study, increased levels of N6-acetyllysine were associated with increased risk of incident stroke among women in the NHS, and incident stroke among men and women in PREDIMED. N6-acetyllysine is an acetylated amino acid that is synthesized from lysine by selective acetylation of the terminal amine group.[36] Acetylation of lysine residues in histone proteins regulates the binding of histones to DNA in nucleosomes, providing a mechanism for the control of the expression of genes on that DNA locus.[37] Previous literature has highlighted the broad role of acetylation of proteins in cell regulation.[38] The large majority of intermediate metabolic

Figure 4 Differential Network of 16 Nodes and 28 Edges Comparing Ischemic Stroke Cases to Controls in the Nurses' Health Study (NHS)

Edges included in the differential network satisfied a false discovery rate *p* value threshold of 0.05. Edge width is proportional to the magnitude of the difference in partial correlation between ischemic stroke cases and controls. Red edges indicate edges with higher partial correlation in the stroke cases relative to controls; black edges indicate edges with higher partial correlation in the controls relative to stroke cases. Node colors denote community membership.

enzymes are acetylated, thereby affecting their enzymatic activity, and the rate of acetylation of metabolic enzymes is dynamically influenced by changes in extracellular conditions.[38] N6-acetyllysine has not been previously implicated in a population-based study of stroke. However, among patients with mild renal disease, higher levels of N6-acetyllysine were associated with increased risk of progression to end stage renal disease, after adjusting for blood pressure, BMI, smoking status, hemoglobin A1C, albumin to creatinine ratio, estimated glomerular filtration rate, uric acid levels, and medication use.[39]

Increased levels of methionine sulfoxide (MetSO) were associated with increased risk of ischemic stroke in NHS and were replicated in PREDIMED. MetSO is a product of oxidative stress. Genetic variation in the MetSO reductase-3 gene (*MSRB3* minor allele G), which converts MetSO back to methionine, was associated with increased odds of brain infarct on MRI in the Framingham Heart Study.[40] In contrast to our findings, one study of men with poorly controlled diabetes mellitus found that lower levels of MetSO were associated with increased risk of CVD[41]; however, we observed significant effect modification by sex for the association of MetSO with incident stroke risk, observing stroke risk in women but not in men.

In our study, increased levels of sucrose/lactose/trehalose were associated with increased risk of incident ischemic stroke

in the NHS, and this association was replicated in PREDIMED. Sucrose/lactose/trehalose are sugars resulting from dietary sugar consumption. Our finding is consistent with the well-established literature on the link between increased dietary sugar consumption and heightened risks of obesity, type 2 diabetes, and CVD, including stroke.[42,43] In the prospective metabolomics study of incident stroke in ARIC (n = 3,904, 114 strokes),[17] 3 sugars (glucose, trehalose, and mannose) measured in serum were positively associated with stroke risk in a model adjusted for age, sex, field center, and batch; however, this association was attenuated after adjustment for traditional risk factors. In the analysis by Vojinovic et al.,[18] glucose was associated with increased risk of incident stroke, but their models did not control for baseline hemoglobin A1C.[18] In our study, the association of sucrose/lactose/ trehalose with stroke was robust even after controlling for cardiovascular risk factors as well as baseline hemoglobin A1C levels.

Increased levels of sebacate were strongly associated with incident stroke in NHS. Sebacate was not measured in PREDIMED and its association with incident stroke risk was close to the threshold for validation in NHS2 but did not reach significance. Sebacate, also known as decanedioate and sebacic acid, is a naturally occurring, saturated, straight-chain dicarboxylic acid. A prospective study of 114 strokes in 3,904 men and women from the ARIC

study identified tetradecanedioate and hexadecanedioate as being strongly associated with incident ischemic stroke after adjusting for risk factors.[17] Both tetradecanedioate and hexadecanedioate have similar chemical structures to sebacate (or decanedioate). Hexadecanedioate has been associated with blood pressure.[44] Further investigation of the underlying pathways and mechanisms resulting in the association of these dicarboxylic acids and stroke is warranted.

Very few metabolites in our study overlapped with those measured in a recent meta-analysis of incident total stroke and 147 serum/plasma metabolites from 7 prospective studies by Vojinovic et al.[18] Of the 10 metabolites with significant stroke associations reported, 2 amino acids, histidine and phenylalanine, were also measured in our platform. However, these 2 amino acids were not significant in our analyses in either minimally or fully adjusted models in the NHS. Differences in the metabolomic platforms may have contributed to differences in the findings. The meta-analysis by Vojinovic et al.[18] measured metabolites using nuclear magnetic resonance (NMR) technology, mostly on the Nightingale platform, compared with our measures by LC-MS.

Our prospective study of the metabolomics of incident ischemic stroke in women fills an important gap in the literature, providing the largest study of the metabolomics of stroke from a single cohort, validating the metabolites in an independent dataset and evaluating metabolites in risk prediction. Strengths include a well-validated LC-MS metabolomics platform, detailed covariate information, and a large number of carefully adjudicated endpoints. Our study employed a robust, 2-stage discovery and validation framework in 3 independent datasets. A limitation of this study was that metabolomics profiling was obtained at a single point in time, resulting in an inability to evaluate longitudinal changes in the metabolites over time in relationship to stroke risk. Moreover, while 95% of the stroke cases in the NHS had imaging with CT or MRI, only 69% of the stroke cases in NHS2 had imaging available. In addition, our discovery dataset in the NHS was predominantly postmenopausal white women. Thus, future research on the generalizability of these findings to the broader population with diverse race/ethnicity and sex compositions is critical.

We identified 23 metabolites associated with incident ischemic stroke in women, and validated 4 of these in independent cohorts. A score comprising the 4 validated metabolites resulted in significant improvement in stroke risk prediction. Further research to identify the mechanisms responsible for these findings, as well as to replicate other metabolites, is needed.

Study Funding

Disclosure

Publication History

Appendix Authors

Name	Location	Contribution
Raji Balasubramanian, ScD	Department of Biostatistics and Epidemiology, University of Massachusetts Amherst	Drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data
Jie Hu, MD, PhD	Division of Women's Health, Brigham and Women's Hospital, Harvard Medical School, Boston, MA	Drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data
Marta Guasch-Ferre, PhD	Department of Nutrition, Harvard T.H. Chan School of Public Health; Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA	Drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data
Jun Li, MD, PhD	Department of Nutrition, Harvard T.H. Chan School of Public Health; Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA	Analysis or interpretation of data
Farzaneh Sorond, MD	Davee Department of Neurology, Division of Stroke and Neurocritical Care, Northwestern Feinberg School of Medicine, Chicago, IL	Drafting/revision of the manuscript for content, including medical writing for content
Yibai Zhao, MS	Department of Biostatistics and Epidemiology, University of Massachusetts Amherst	Analysis or interpretation of data
Katherine H. Shutta, MS	Department of Biostatistics and Epidemiology, University of Massachusetts Amherst	Drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data
Jordi Salas-Salvadó, MD, PhD	Universitat Rovira i Virgili, Departament de Bioquímica i Biotecnologia, Unitat de Nutrició, Reus; Centro de Investigación Biomédica en Red Fisiopatología de la Obesidad y la Nutrición (CIBEROBN), Institute of Health Carlos III, Madrid; Pere Virgili Research Institute (IISPV), University Hospital of Sant Joan de Reus, Nutrition Unit, Reus, Spain	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data

Appendix *(continued)*

Name	Location	Contribution
Frank Hu, MD, PhD	Department of Nutrition, Harvard T.H. Chan School of Public Health; Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School; Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Clary B. Clish, PhD	Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; analysis or interpretation of data
Kathryn M. Rexrode, MD, MPH	Division of Women's Health, Brigham and Women's Hospital, Harvard Medical School, Boston, MA	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation of data

References

1. Virani SS, Alonso A, Benjamin EJ, et al. Heart disease and stroke statistics-2020 update: a report from the American Heart Association. *Circulation.* 2020;141(9): e139-e596.
2. Carandang R, Seshadri S, Beiser A, et al. Trends in incidence, lifetime risk, severity, and 30-day mortality of stroke over the past 50 years. *JAMA.* 2006;296(24): 2939-2946.
3. Peters SA, Huxley RR, Woodward M. Diabetes as a risk factor for stroke in women compared with men: a systematic review and meta-analysis of 64 cohorts, including 775 385 individuals and 12 539 strokes. *Lancet.* 2014;383(9933):1973–1980.
4. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. *Nat Med.* 2011;17(4):448-453.
5. Wang Z, Klipfell E, Bennett BJ, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature.* 2011;472(7341):57-63.
6. Iso H, Rexrode KM, Stampfer MJ, et al. Intake of fish and omega-3 fatty acids and risk of stroke in women. *JAMA.* 2001;285(3):304-312.
7. Larsson SC, Orsini N, Wolk A. Long-chain omega-3 polyunsaturated fatty acids and risk of stroke: a meta-analysis. *Eur J Epidemiol.* 2012;27(12):895-901.
8. Würtz P, Havulinna AS, Soininen P, et al. Metabolite profiling and cardiovascular event risk: a prospective study of three population-based cohorts. *Circulation.* 2015; 114:013116.
9. Wang Z, Klipfell E, Bennett BJ, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature.* 2011;472(7341):57-63.
10. Vaarhorst AA, Verhoeven A, Weller CM, et al. A metabolomic profile is associated with the risk of incident coronary heart disease. *Am Heart J.* 2014;168(1):45-52.e47.
11. Stegemann C, Drozdov I, Shalhoub J, et al. Comparative lipidomics profiling of human atherosclerotic plaques. *Circ Cardiovasc Genet.* 2011;4(3):232-242.
12. Shah SH, Sun J-L, Stevens RD, et al. Baseline metabolomic profiles predict cardiovascular events in patients at risk for coronary artery disease. *Am Heart J.* 2012;163(5): 844-850. e841.
13. Shah SH, Bain JR, Muehlbauer MJ, et al. Association of a peripheral blood metabolic profile with coronary artery disease and risk of subsequent cardiovascular events. *Circ Cardiovasc Genetics.* 2010;109:852814.
14. Ganna A, Salihovic S, Sundström J, et al. Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease. *Plos Genet.* 2014;10(12): e1004801.
15. Cheng S, Larson MG, McCabe EL, et al. Distinct metabolomic signatures are associated with longevity in humans. *Nat Commun.* 2015;6:6791.
16. Paynter NP, Balasubramanian R, Giulianini F, et al. Metabolic predictors of incident coronary heart disease in women. *Circulation.* 2018;137(8):841-853.
17. Sun D, Tiedt S, Yu B, et al. A prospective study of serum metabolites and risk of ischemic stroke. *Neurology.* 2019;92(16):e1890–e1898.
18. Vojinovic D, Kalaoja M, Trompet S, et al. Association of circulating metabolites in plasma or serum and risk of stroke: meta-analysis from seven prospective cohorts. *Neurology.* 2020;96(8):e1110–e1123.
19. Krumsiek J, Mittelstrass K, Do KT, et al. Gender-specific pathway differences in the human serum metabolome. *Metabolomics.* 2015;11(6):1815-1833.
20. Estruch R, Ros E, Martinez-Gonzalez M. Mediterranean diet for primary prevention of cardiovascular disease. *N Engl J Med.* 2013;369(7):676-677.
21. Martínez-González MÁ, Corella D, Salas-Salvadó J, et al. Cohort profile: design and methods of the PREDIMED study. *Int J Epidemiol.* 2012;41(2):377-385.
22. Hankinson SE, Willett WC, Manson JE, et al. Alcohol, height, and adiposity in relation to estrogen and prolactin levels in postmenopausal women. *J Natl Cancer Inst.* 1995; 87(17):1297-1302.
23. Zhang X, Tworoger SS, Eliassen AH, Hankinson SE. Postmenopausal plasma sex hormone levels and breast cancer risk over 20 years of follow-up. *Breast Cancer Res Treat.* 2013;137(3):883-892.
24. Tworoger SS, Eliassen AH, Zhang X, et al. A 20-year prospective study of plasma prolactin as a risk marker of breast cancer development. *Cancer Res.* 2013;73(15): 4810-4819.
25. Walker AE, Robins M, Weinfeld FD. The National Survey of Stroke: clinical findings. *Stroke.* 1981;12(2 Pt 2 suppl 1):I13-I44.
26. Iso H, Rexrode K, Hennekens CH, Manson JE. Application of computer tomography-oriented criteria for stroke subtype classification in a prospective study. *Ann Epidemiol.* 2000;10(2):81-87.
27. Estruch R, Ros E, Salas-Salvado J, et al. Primary prevention of cardiovascular disease with a Mediterranean diet. *N Engl J Med.* 2013;368(14):1279-1290.
28. Ruiz-Canela M, Toledo E, Clish CB, et al. Plasma branched-chain amino acids and incident cardiovascular disease in the PREDIMED trial. *Clin Chem.* 2016;62(4):582-592.
29. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A.* 2003;100(16):9440-9445.
30. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44(3):837-845.
31. Ha MJ, Baladandayuthapani V, Do KA. DINGO: differential network analysis in genomics. *Bioinformatics.* 2015;31(21):3413-3420.
32. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal.* 2006:1695.
33. Nurses' Health Study. Published 2016. Available at: nurseshealthstudy.org. Accessed 2021.
34. Kleinberg JM. Authoritative sources in a hyperlinked environment. *J ACM.* 1999; 46(5):604-632.
35. Freeman LC. Centrality in social networks conceptual clarification. *Social Networks.* 1978;1(3):215-239.
36. Kikugawa Y, Mitsui K, Sakamoto T, Kawase M, Tamiya H. N-methoxydiacetamide: a new selective acetylating agent. *Tetrahedron Lett.* 1990;31(2):243-246.
37. Crane-Robinson C, Hebbes TR, Clayton AL, Thorne AW. Chromosomal mapping of core histone acetylation by immunoselection. *Methods.* 1997;12(1):48-56.
38. Zhao S, Xu W, Jiang W, et al. Regulation of cellular metabolism by protein lysine acetylation. *Science.* 2010;327(5968):1000-1004.
39. Niewczas MA, Mathew AV, Croall S, et al. Circulating modified metabolites and a risk of ESRD in patients with type 1 diabetes and chronic kidney disease. *Diabetes Care.* 2017;40(3):383-390.
40. Conner SC, Benayoun L, Himali JJ, et al. Methionine sulfoxide reductase-B3 risk allele implicated in Alzheimer's disease associates with increased odds for brain infarcts. *J Alzheimers Dis.* 2019;68(1):357-365.
41. Koska J, Saremi A, Howell S, et al. Advanced glycation end products, oxidation products, and incident cardiovascular events in patients with type 2 diabetes. *Diabetes Care.* 2018;41(3):570-576.
42. Bernstein AM, de Koning L, Flint AJ, Rexrode KM, Willett WC. Soda consumption and the risk of stroke in men and women. *Am J Clin Nutr.* 2012;95(5):1190-1199.
43. Fung TT, Malik V, Rexrode KM, Manson JE, Willett WC, Hu FB. Sweetened beverage consumption and risk of coronary heart disease in women. *Am J Clin Nutr.* 2009; 89(4):1037-1042.
44. Menni C, Graham D, Kastenmuller G, et al. Metabolomic identification of a novel pathway of blood pressure regulation involving hexadecanedioate. *Hypertension.* 2015; 66(2):422-429.