

A prospective study of serum metabolites and risk of ischemic stroke

Daokun Sun, MD, MPH,* Steffen Tiedt, MD, PhD,* Bing Yu, PhD, Xueqiu Jian, PhD, Rebecca F. Gottesman, MD, PhD, Thomas H. Mosley, PhD, Eric Boerwinkle, PhD, Martin Dichgans, MD, and Myriam Fornage, PhD

Correspondence

Dr. Fornage
Myriam.Fornage@uth.tmc.edu

Neurology® 2019;92:e1890-e1898. doi:10.1212/WNL.0000000000007279

Abstract

Objective

To identify promising blood-based biomarkers and novel etiologic pathways of disease risk, we applied an untargeted serum metabolomics profiling in a community-based prospective study of ischemic stroke (IS).

Methods

In 3,904 men and women from the Atherosclerosis Risk In Communities study, Cox proportional hazard models were used to estimate the association of incident IS with the standardized level of 245 fasting serum metabolites individually, adjusting for age, sex, race, field center, batch, diabetes, hypertension, current smoking status, body mass index, and estimated glomerular filtration rate. Validation of results was carried out in an independent sample of 114 IS cases and 112 healthy controls.

Results

Serum levels of 2 long-chain dicarboxylic acids, tetradecanedioate and hexadecanedioate, were strongly correlated ($r = 0.88$) and were associated with incident IS after adjusting for covariates (hazard ratio [95% confidence interval (CI)] 1.11 [1.06–1.16] and 1.12 [1.07–1.17], respectively; $p < 0.0001$). Analyses by IS subtypes suggested that these associations were specific to cardioembolic stroke (CES). Associations of tetradecanedioate and hexadecanedioate with IS were independently confirmed (odds ratio [95% CI] 1.76 [1.21; 2.56] and 1.60 [1.11; 2.32], respectively).

Conclusion

Two serum long-chain dicarboxylic acids, metabolic products of ω -oxidation of fatty acids, were associated with IS and CES independently of known risk factors. Pathways related to intracellular hexadecanedioate synthesis or those involved in its clearance from the circulation may mediate IS risk. These results highlight the potential of metabolomics to discover novel circulating biomarkers for stroke and to unravel novel pathways for IS and its subtypes.

*These authors contributed equally to this work.

From the Brown Foundation Institute of Molecular Medicine, McGovern Medical School (D.S., X.J., M.F.), and School of Public Health (B.Y., E.B., M.F.), The University of Texas Health Science Center at Houston; Institute for Stroke and Dementia Research (S.T., M.D.), Klinikum der Universität München, Ludwig-Maximilians-Universität LMU, Munich, Germany; Johns Hopkins University School of Medicine (R.F.G.), Baltimore, MD; The University of Mississippi Medical Center (T.H.M.), Jackson; German Center for Neurodegenerative Diseases (DZNE, Munich) (M.D.); and Munich Cluster for Systems Neurology (SyNergy) (S.T., M.D.), Germany.

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.

The Article Processing Charge was funded by NIH.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND), which permits downloading and sharing the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Glossary

AF = atrial fibrillation; **ARIC** = Atherosclerosis Risk in Communities; **BP** = blood pressure; **CES** = cardioembolic stroke; **CI** = confidence interval; **CIRCULAS** = Circulating Biomarkers in Acute Stroke; **eGFR** = estimated glomerular filtration rate; **GWAS** = genome-wide association study; **HC** = healthy control; **HR** = hazard ratio; **ICD-9-CM** = International Classification of Diseases–9–clinical modification; **IS** = ischemic stroke; **KUM** = Klinikum der Universität München; **LAA** = large artery atherosclerosis; **MI** = myocardial infarction; **SNP** = single nucleotide polymorphism.

Stroke is a leading cause of death and disability worldwide. According to the WHO, 15 million individuals have a stroke each year. Five million of them die and another 5 million are left disabled.¹ Ischemic strokes (IS) account for approximately 80% of all strokes. Despite decades of research, few nonpalliative treatments are available for IS. IV administration of a thrombolytic agent shortly after occurrence of the acute event is the only approved treatment to re-establish blood flow.² Management of stroke risk factors remains the first line strategy to decrease the burden of IS. In addition, stroke classification and diagnosis are not straightforward and mostly rely on neuroimaging techniques that are costly, time-consuming, and not universally available.³ Thus, the discovery of novel blood-based biomarkers that aid in early identification of high-risk patients and in classification and diagnosis of stroke has potential for facilitating management of patients with cerebral ischemia and improving understanding of IS etiology.

The human blood contains a wide variety of chemically diverse low molecular weight compounds, the metabolome, which can be measured in parallel through modern metabolomic technologies. Analysis of the metabolome provides new opportunities in the investigation of stroke biomarkers. Recent studies have identified altered serum and urine metabolic profiles in patients with cerebral infarction.^{4,5} However, these studies were limited by their cross-sectional study design. Here, we conducted a prospective analysis of the relationship between serum metabolites and incident IS in 3,904 middle-aged adults followed up for a median of 26 years.

Methods

Study sample

The Atherosclerosis Risk in Communities (ARIC) study is a prospective, biracial population-based study of atherosclerosis and cardiovascular diseases in 15,792 adults (11,478 non-Hispanic white; 8,710 women) aged between 45 and 64 years at the baseline examination in 1987–1989. Participants were randomly selected and recruited from communities in suburban Minneapolis, Minnesota; Washington County, Maryland; and Forsyth County, North Carolina. In Jackson, Mississippi, only black residents were enrolled. Details about the ARIC study design and examination procedures have been published.⁶ Metabolomics profiling was performed in a subsample of 4,032 participants (61.5% African Americans) who provided informed consent and had a stored fasting

(≥8 hours) serum sample at the baseline examination. Only individuals free of stroke or TIA at baseline were included in the analyses (n = 3,904).

Stroke ascertainment

The study included hospitalized strokes that occurred by December 31, 2015. Participants' hospitalizations and deaths were identified via annual telephone contacts and by review of hospital records and lists of stroke discharges (ICD-9-CM codes 430–438). Details on quality assurance for ascertainment and classification of stroke are described elsewhere.^{7,8} Briefly, stroke diagnosis was assigned via a computer algorithm and an expert reviewer, based on criteria adapted from the National Survey of Stroke.⁹ Strokes were classified as hemorrhagic stroke (subarachnoid and intracerebral hemorrhage) and IS (thrombotic and embolic brain infarction) based on neuroimaging, autopsy, and review of the medical records. All definite IS were further classified as small vessel disease, large artery atherosclerosis (LAA), or cardioembolic stroke (CES) on the basis of the recorded neuroimaging results and the review of the medical records. For this analysis, hemorrhagic strokes identified by ARIC were censored at the time of occurrence.

Covariate assessment

Baseline resting blood pressure (BP) and weight were measured following standardized protocols. Hypertension was defined as taking antihypertensive medication or having an untreated BP of ≥140 mm Hg (systolic) or ≥90 mm Hg (diastolic). Antihypertensive use during the last 2 weeks was ascertained by inventory.

Body mass index was measured as weight in kilograms divided by the square of height in meters. Smoking status was self-reported. Diabetes was defined as having fasting blood glucose ≥126 mg/dL or receiving antidiabetic therapies. Prevalent coronary heart disease was defined as evidence of a previous myocardial infarction (MI) by ECG at the baseline examination, a history of physician-diagnosed MI, or a previous cardiac revascularization procedure. Estimated glomerular filtration rate (eGFR) was calculated using the Chronic Kidney Disease Epidemiology Collaboration equation.¹⁰

Metabolomics

Metabolomic profiling was performed by Metabolon Inc. (Durham, NC) using an untargeted, gas chromatography–mass spectrometry and liquid chromatography–mass spectrometry–based quantification protocol.^{11,12} Metabolite

levels were measured in fasting serum samples collected at the baseline examination (1987–1989) and processed using Metabolon's standard solvent extraction method. Instrument variability was determined by calculating the median relative SD for internal standards added to each sample prior to analysis. Overall process variability was determined by calculating the median relative SD for endogenous metabolites present in all of the technical replicate samples. Moreover, data normalization was performed to correct for variation resulting from day-to-day instrument tuning differences. Metabolites were identified by automated spectral comparison to a standard library and missing values were imputed using minimum nonmissing measurement. In ARIC, metabolite profiles were measured in 2 phases or batches (2010 and 2014). A total of 384 metabolites (named compounds) that overlapped the 2 batches were detected and quantified. Of these, 285 had $\leq 25\%$ missing values. We also designed a set of 97 baseline serum samples and measured their metabolome profiles in 2010 and 2014. We calculated the Pearson correlation coefficients (r) between the 97 pairs for these 285 metabolites. Among 285 metabolites, the correlation ranged from -0.09 to 0.99 , with mean of 0.63 and median of 0.71 . We excluded 40 metabolites with poor correlation coefficient ($r < 0.3$). Therefore, 245 metabolites with good long-term stability and reliability were examined in this study.

Statistical analysis

Baseline characteristics of study participants with and without incident IS were compared using χ^2 and t tests for categorical and continuous variables, respectively. We analyzed 245 metabolites that had values above the detection limit in at least 75% of the samples. For each metabolite, standardized values (centered by mean and scaled by SD) were derived and used in the analyses. Cox proportional hazards regression models using age as the time scale were estimated to assess the association of each metabolite with incident IS and subtypes adjusting for covariates. Two models were used that adjusted for age, sex, race, field center, and batch (model 1); and additionally for diabetes, hypertension, current smoking, body mass index, and eGFR (model 2). Results are reported as hazard ratios per SD increase in standardized serum metabolite. A modified stepwise Bonferroni procedure, the Dubey and Armitage-Parmer algorithm, was used to correct for multiple testing.¹³ This method takes into account the correlation structure of the metabolites. Accordingly, a p value $< 3.0 \times 10^{-4}$ was considered statistically significant. A false discovery rate was also estimated.¹⁴ All analyses were performed with SAS version 9.4 (SAS Institute, Cary, NC). We also checked the proportional hazards assumption for the fitted Cox models for the most significantly associated metabolites by calculating the Schoenfeld residuals and implementing graphical diagnostics using the *survival* and *survminer* R packages.

Validation of findings in the Circulating Biomarkers in Acute Stroke (CIRCULAS) study

Results were validated in an independent sample of IS cases and controls recruited from the Klinikum der Universität

München (KUM), a tertiary level hospital at Ludwig-Maximilians-Universität, Munich, Germany. Recruitment procedures have been described.¹⁵ Briefly, patients with suspected IS within 24 hours of symptom onset were prospectively recruited from the emergency department through the CIRCULAS study. All patients had a final diagnosis of IS as defined by an acute focal neurologic deficit in combination with a diffusion-weighted imaging–positive lesion on MRI, or a new lesion on a delayed CT scan.¹⁶ The sample consisted of 2 subsamples, one including 74 IS cases and 73 healthy controls (HCs) enrolled between February 2014 and August 2015 and the other an additional 40 IS cases and 40 HCs enrolled between September 2015 and November 2016. HCs from the latter subsample were matched to patients with IS for age, sex, hypertension, smoking history, hypercholesterolemia, obesity, diabetes mellitus, and family history.

Serum samples from IS patients were collected immediately upon hospital arrival in the emergency department. Samples from HCs were collected in the outpatient clinic of KUM. Serum metabolites were measured by Metabolon as described above. We excluded 1 HC, who showed missing values for $\geq 10\%$ of the metabolites. For this study, we performed a targeted look-up and exclusively extracted relative concentrations of hexadecanedioate and tetradecanedioate from the comprehensive metabolomic analysis in CIRCULAS patients. For each metabolite, standardized values (centered by mean and scaled by SD) were derived.

We used multivariable logistic regression to evaluate association of each metabolite with IS status. Two models were used that adjusted for baseline age, sex (model 1), and additionally for platelet count, glucose, body mass index, and eGFR (model 2). All analyses were performed in R, version 3.5.0.¹⁷ Analyses were performed separately in the 2 subsamples and combined by fixed-effects inverse variance weighted meta-analysis.

Genetic association of selected single nucleotide polymorphisms (SNPs) with incident ischemic and CES

Six missense variants previously reported to be associated with serum hexadecanedioate levels ($p < 5 \times 10^{-8}$) in a genome-wide association study¹⁸ were examined for their association with IS and CES in the largest genetic study of stroke to date (the MEGASTROKE study).¹⁹ A Bonferroni correction accounting for the number of SNPs tested was applied to determine the p value threshold of statistical significance ($p = 0.05/6 [0.008]$).

Standard protocol approvals, registrations, and patient consents

The ARIC study has been approved by the institutional review boards at the University of Minnesota, Johns Hopkins University, Wake Forest University, the University of North Carolina–Chapel Hill, the University of Texas Health Science Center at Houston, and the University of Mississippi Medical

Center. All ARIC participants included in this research provided written informed consent. The CIRCULAS study was conducted in accordance with the Declaration of Helsinki as well as institutional guidelines. Written informed consent was obtained from all participants in conformity with the approval by the local ethics committee.

Data availability

The data used in this manuscript can be obtained through the ARIC Coordinating Center at the University of North Carolina–Chapel Hill. Procedures for data request can be found on the following website: www2.csc.unc.edu/aric/distribution-agreements.

Results

A total of 3,904 ARIC participants who had metabolomics profiling and were free of stroke at the baseline examination were included in this study. This subset was oversampled on African American race (61% African American participants in this subset vs ~26% in the overall cohort). Compared to those without metabolomics profiling, participants in this study were generally younger, more likely to be women, had higher rates of hypertension and diabetes, but similar rates of coronary heart disease and smoking (table e-1: doi.org/10.5061/dryad.55hr85m). In this subset, 346 incident IS cases were observed

during a median follow-up of 26 years. Table 1 shows selected baseline characteristics of the sample by IS status during follow-up. Compared to those without incident IS diagnosis, those who developed an IS were older, more likely to be male, African American, and current smoker, had higher BP, and higher prevalence of diabetes, hypertension, and coronary heart disease.

In minimally adjusted models, 9 metabolites were associated with incident IS (table 2). After further adjustment for known IS risk factors and kidney function, 2 metabolites remained associated: tetradecanedioate (hazard ratio [HR] per SD higher level: 1.11 [95% confidence interval (CI) 1.06–1.16]) and hexadecanedioate (HR [95% CI] 1.12 [1.07–1.17]). These metabolites were strongly correlated ($r = 0.88$). As expected, when included simultaneously in the Cox model, associations of both metabolites were weaker but remained nominally significant for hexadecanedioate (model 1: HR [95% CI] 1.22 [1.04; 1.44], $p = 0.02$; model 2: HR [95% CI] 1.14 [1.00; 1.39], $p = 0.09$) but not tetradecanedioate (model 1: HR [95% CI] 0.93 [0.80–1.08], $p = 0.35$; model 2: HR [95% CI] 0.98 [0.81; 1.17], $p = 0.80$). Association of the 2 metabolites with IS subtypes is shown in table 3. Hexadecanedioate was associated with incident CES (model 1: HR [95% CI] 1.14 [1.07; 1.23], $p = 0.0003$; model 2: HR [95% CI] 1.14 [1.06; 1.24], $p = 0.0007$) but not with other

Table 1 Selected baseline characteristics of the study sample by ischemic stroke status during follow-up

Variable	Ischemic stroke	No ischemic stroke	<i>p</i> Value
N	346	3,558	
Age, y, mean (SD)	55.1 (5.9)	53.4 (5.7)	<0.0001
Women, %	52.0	61.0	0.0014
Black, %	67.3	60.7	0.01
Prevalent CHD, %	8.2	4.2	0.0007
Current smoking, %	31.8	26.6	0.0038
Diabetes, %	22.0	10.4	<0.0001
Hypertension, %	58.9	42.6	<0.0001
HBP Rx use, %	46.8	34.3	<0.0001
eGFR (creatinine), mL/min/1.73 m ² , mean (SD)	105.4 (19.2)	106.8 (18.6)	0.19
Systolic blood pressure, mm Hg, mean (SD)	132.6 (21.9)	123.89 (20.5)	<0.0001
Diastolic blood pressure, mm Hg, mean (SD)	80.3 (13.2)	76.3 (12.1)	<0.0001
Body mass index, kg/m ² , mean (SD)	29.1 (5.5)	28.7 (5.8)	0.29
Follow-up time, y, mean (SD)	14.6 (7.1)	22.5 (7.1)	<0.0001
LAA, n	24	NA	
SVD, n	67	NA	
CES, n	94	NA	

Abbreviations: CES = cardioembolic stroke; CHD = coronary heart disease; eGFR = estimated glomerular filtration rate; HBP Rx = antihypertensive medication; LAA = large artery atherosclerosis; SVD = small vessel disease.

Table 2 Association of serum metabolites with incident ischemic stroke in a subsample of the Atherosclerosis Risk in Communities study, 1987–2015

Metabolite	Model 1			Model 2		
	HR (95% CI)	p Value	FDR	HR (95% CI)	p Value	FDR
Glycine	0.78 (0.64; 0.91)	2.0×10^{-4}	1.3×10^{-2}	0.86 (0.73; 0.99)	0.02	0.47
DSGEGDFXAEGGGVR (fibrinogen cleavage peptide)	1.10 (1.05; 1.15)	3.0×10^{-4}	1.6×10^{-2}	1.10 (1.04; 1.15)	8.0×10^{-4}	0.59
Glucose	1.29 (1.21; 1.37)	4.1×10^{-10}	3.8×10^{-6}	1.10 (0.99; 1.22)	0.09	0.56
Trehalose	1.17 (1.11; 1.24)	1.2×10^{-6}	3.0×10^{-4}	1.07 (0.98; 1.16)	0.13	0.59
Mannose	1.25 (1.17; 1.32)	1.7×10^{-9}	3.8×10^{-6}	1.11 (0.99; 1.20)	0.08	0.55
Tetradecanedioate	1.11 (1.07; 1.16)	3.7×10^{-6}	8.4×10^{-4}	1.11 (1.06; 1.16)	4.9×10^{-5}	1.6×10^{-2}
Hexadecanedioate	1.13 (1.08; 1.19)	2.4×10^{-7}	6.6×10^{-5}	1.12 (1.07; 1.17)	3.2×10^{-5}	1.2×10^{-2}
Arabonate	1.13 (1.07; 1.20)	3.0×10^{-4}	1.6×10^{-2}	1.05 (0.96; 1.14)	0.279	0.69

Abbreviations: CI = confidence interval; FDR = false discovery rate; HR = hazard ratio.

Model 1: Cox proportional hazards model adjusted for age, sex, race, field center, and batch; model 2: same as model 1 additionally adjusted for diabetes, hypertension, current smoking, body mass index, and estimated glomerular filtration rate.

stroke subtypes. Consistent results were observed for tetradecanedioate but these reached only nominal significance (model 1: HR [95% CI] 1.11 [1.03; 1.21], $p = 0.01$; model 2: HR [95% CI] 1.11 [1.02; 1.21], $p = 0.02$). Using Schoenfeld residuals tests, we found no evidence of violation of the proportional hazards assumption in Cox models estimating the association of hexadecanedioate with incident CES (global test: model 1 $p = 0.34$; model 2 $p = 0.75$; individual test: model 1 $p = 0.83$; model 2 $p = 0.79$).

A previous study reported that use of statins was associated with lower levels of hexadecanedioate.²⁰ This was not observed in ARIC ($p = 0.94$). Nonetheless, we repeated our analyses further adjusting for statin use at baseline. The association with incident CES remained unchanged (model 1:

HR [95% CI] 1.14 [1.07; 1.23], $p = 0.0003$; model 2: HR [95% CI] 1.14 [1.05; 1.24], $p = 0.0008$).

Atrial fibrillation (AF) and CES are strongly associated.²¹ Thus, we examined whether hexadecanedioate levels showed evidence of association with incident AF in 3,922 ARIC participants (608 cases) as previously reported.²² Three models were estimated: model 1 adjusted for age, sex, field center, race, and batch; model 2 additionally adjusted for smoking status, body mass index, systolic BP, hypertension medications, diabetes, prevalent heart failure, and prevalent coronary heart disease; model 3 further adjusted for eGFR. Consistent with the observed association with incident CES, hexadecanedioate levels were associated with incident AF (HR [95% CI] 1.11 [1.04, 1.18], $p = 0.004$ [model 1]; 1.08

Table 3 Association of the 2 serum metabolites with incident ischemic stroke subtypes in a subsample of the Atherosclerosis Risk in Communities study, 1987–2015

Metabolite	Model 1		Model 2	
	HR (95% CI)	p Value	HR (95% CI)	p Value
Tetradecanedioate				
LAA	1.17 (0.98; 1.40)	0.08	1.17 (0.96; 1.42)	0.12
SVD	1.01 (0.71; 1.43)	0.95	0.96 (0.65; 1.41)	0.82
CES	1.11 (1.03; 1.21)	0.01	1.11 (1.02; 1.21)	0.02
Hexadecanedioate				
LAA	1.15 (0.87; 1.52)	0.31	1.13 (0.80; 1.59)	0.49
SVD	1.06 (0.82; 1.36)	0.66	0.97 (0.70; 1.34)	0.84
CES	1.14 (1.07; 1.23)	0.0003	1.14 (1.06; 1.24)	0.0007

Abbreviations: CES = cardioembolic stroke; CI = confidence interval; HR = hazard ratio; LAA = large artery atherosclerosis; SVD = small vessel disease.

Model 1: Cox proportional hazards model adjusted for age, sex, race, field center, and batch; model 2: same as model 1 additionally adjusted for diabetes, hypertension, current smoking, body mass index, and estimated glomerular filtration rate.

Table 4 Association of the 2 serum metabolites with ischemic stroke in the Circulating Biomarkers in Acute Stroke study

Metabolite	Model 1		Model 2	
	OR (95% CI)	p Value	OR (95% CI)	p Value
Tetradecanedioate	1.76 (1.30; 2.40)	0.0002	1.76 (1.21; 2.56)	0.003
Hexadecanedioate	1.72 (1.27; 2.32)	0.0004	1.60 (1.11; 2.32)	0.012

Abbreviations: CI = confidence interval; OR = odds ratio.

Model 1: adjusted for age and sex; model 2: same as model 1 additionally adjusted for platelet count, glucose levels, body mass index, and estimated glomerular filtration rate.

[1.01, 1.15], $p = 0.049$ [model 2]; 1.08 [1.00, 1.15], $p = 0.05$ [model 3]).

We sought to independently validate our findings in the CIRCULAS sample comprising a total of 114 IS cases and 112 HCs recruited from the KUM. Demographic characteristics of the CIRCULAS sample are available in table e-2 (doi.org/10.5061/dryad.55hr85m). As shown in table 4, both tetradecanedioate and hexadecanedioate were associated with acute IS (odds ratio [95% CI] 1.76 [1.21, 2.56], $p = 0.003$, and 1.60 [1.11, 2.32], $p = 0.012$, respectively).

A genome-wide association study (GWAS)¹⁸ had previously identified regions of the human genome harboring DNA variants associated with serum hexadecanedioate levels in the TwinsUK and KORA cohorts of adults of European ancestry. We therefore examined the association between these genetic variants and risk of IS and CES in the largest GWAS published to date.¹⁹ We specifically focused on 6 predicted deleterious missense variants that reached genome-wide significance ($p < 5 \times 10^{-8}$) in the previously reported GWAS of

hexadecanedioate.¹⁸ Consistent with the previous report, these variants were strongly associated with serum hexadecanedioate levels in the ARIC participants (table 5). rs11045819, which encodes a proline to threonine substitution at codon 155 of the solute carrier organic anion transporter family member 1B1 (OATP1B1), was significantly associated with CES and nominally with IS. The common allele of this variant was associated with greater serum hexadecanedioate levels ($p = 1.4 \times 10^{-12}$) but with a 8% lower risk of incident CES ($p = 0.003$) and 3% lower risk of IS ($p = 0.04$). Other variants in moderate linkage disequilibrium with rs11045819 showed similar results (table 5).

Discussion

This study reports on the prospective association of untar-geted serum metabolomics profiles with development of IS over a 26-year period in a community-based sample of older adults of European and African American ancestry. We identified 2 long-chain dicarboxylic acids, tetradecanedioate and hexadecanedioate, whose serum levels were highly

Table 5 Associations of selected genetic variants previously reported to influence circulating hexadecanedioate levels with ischemic stroke (IS) and cardioembolic stroke (CES) in the MEGASTROKE project

SNP	Coded allele	CAF	Gene	Amino acid change	LD with rs11045819	Association with CES (MEGASTROKE)		Association with IS (MEGASTROKE)		Association with hexadecanedioate levels (ARIC)	
						OR (95% CI)	p Value	OR (95% CI)	p Value	β (SE)	p Value
rs4646487	C	0.87	CYP4B1	R173W	—	1.00 (0.94, 1.05)	0.89	1.03 (1.00, 1.06)	0.10	0.07 (0.02)	1.3×10^{-3}
rs11045699	T	0.80	SLCO1B7	F454L	0.76	0.93 (0.89, 0.97)	0.002	0.93 (0.89, 0.97)	0.11	0.08 (0.02)	1.6×10^{-6}
rs11045681	A	0.80	SLCO1B7	Y311C	0.76	0.94 (0.89, 0.98)	0.009	0.98 (0.96, 1.01)	0.13	0.08 (0.02)	1.4×10^{-6}
rs10841795	A	0.86	SLCO1A2	I13T	0.64	0.92 (0.87, 0.98)	0.006	0.96 (0.93, 0.99)	0.013	0.11 (0.02)	4.9×10^{-8}
rs4149056	C	0.16	SLCO1B1	V174A	0.04	0.98 (0.93, 1.04)	0.53	0.96 (0.94, 0.99)	0.01	0.14 (0.02)	2.3×10^{-15}
rs11045819	C	0.84	SLCO1B1	P155T	1	0.92 (0.88, 0.97)	0.003	0.97 (0.95, 1.0)	0.04	0.12 (0.02)	1.4×10^{-12}

Abbreviations: ARIC = Atherosclerosis Risk in Communities; CAF = coded allele frequency; CI = confidence interval; LD = linkage disequilibrium; OR = odds ratio; SNP = single nucleotide polymorphism.

Associations of these SNPs with hexadecanedioate levels in ARIC are also shown.

correlated and were associated with incident IS independently of known risk factors. Analyses by IS subtype revealed associations with CES. In addition, higher levels of these metabolites were also associated with an increased risk of IS in an independent sample of patients with acute IS compared to healthy controls.

These compounds are metabolic products of fatty acids produced by ω -oxidation in the endoplasmic reticulum, a normally minor catabolic pathway, which becomes more prominent when peroxisomal or mitochondrial β -oxidation is defective.²³ A recent study comparing the metabolomics profiles of carotid atherosclerotic plaques extracted from 5 patients manifesting cerebrovascular symptoms of carotid origin with those from 5 asymptomatic patients showed evidence of a downregulation of the β -oxidation pathway in the symptomatic patients.²⁴ Our results showing increased serum levels of several metabolic products of the rescue ω -oxidation pathway in association with the development of CES is consistent with such findings. Moreover, the ω -oxidation pathway has been shown to be active in brain. Indeed, cultured rat neurons, astrocytes, and oligodendrocytes all contain ω -oxidation activity, the product of which is almost exclusively dicarboxylic acid.²⁵

Hexadecanedioate has recently been associated with BP levels and all-cause mortality in Europeans.²⁶ Experimental work in rat models of hypertension further showed that oral administration of hexadecanedioate increases circulating levels of the metabolite and raises BP in Wistar Kyoto rats. Moreover, the mesenteric resistance artery contractile response to noradrenaline was increased in hexadecanedioate-treated rats, suggesting a vascular mechanism in the modulation of BP.²⁶ In our analyses, adjusting for BP did not mitigate the associations of hexadecanedioate with incident IS or CES, suggesting that the relationship between hexadecanedioate and BP cannot fully explain the observed associations.

Studies in experimental models have shed light on possible hexadecanedioate physiologic functions, including its hypolipidemic, anti-obesity, and anti-diabetogenic properties.²⁷ It was also shown to uncouple oxidative phosphorylation in liver cells and in isolated liver and heart mitochondria.²⁸ Whether and how these functions are relevant to IS and, more specifically, CES risk remains to be addressed. Diabetes, dyslipidemia, and obesity are known risk factors for IS.²⁹ Mitochondrial dysfunction has been implicated in several cardiac diseases,³⁰ and reactive oxygen species generated from mitochondria play a major role in the development of AF, a major risk factor for CES.³¹ Interestingly, elevated serum hexadecanedioate was also nominally associated with an increased risk of incident atrial fibrillation in ARIC.

Genetic determinants of circulating hexadecanedioate levels may help shed light on the molecular pathways underlying the relationships between this metabolite and CES. Missense variants mapping to 2 regions previously identified in a GWAS of circulating hexadecanedioate levels¹⁸ were also associated

with serum levels of the metabolites in ARIC but their association with incident IS and CES was ambiguous. Variants in genes encoding solute carrier organic anion transporters, including *SLCO1B1* at chr12p12.1, tended to be inversely associated with CES and IS. *SLCO1B1* encodes OATP1B1, a highly expressed organic anion transporter in the human liver and responsible for the hepatic uptake of drugs and endogenous compounds. In vitro studies recently identified hexadecanedioate and tetradecanedioate as both inhibitors and substrates of OATP1B1.³² Thus, genetic variation at the *SLCO1B1* locus may reflect circulating hexadecanedioate levels as determined by hepatic uptake of the metabolite by OATP1B1, but may have little influence on intracellular levels of the compound. Hence, the inverse association of the *SLCO1B1* locus with IS and CES is consistent with the hypothesis that circulating hexadecanedioate may be a marker of an unrecognized etiologic pathway, possibly related to the clearance function of the transporter. The first step of ω -oxidation of fatty acids that produces hexadecanedioate is carried out by cytochrome P450 enzymes mainly belonging to the CYP4 family.³³ Association of variants at the 1p33 locus with both higher hexadecanedioate levels and an increased risk of IS would be consistent with the hypothesis of a role of intracellular hexadecanedioate in stroke etiology. Although the assessed variant in *CYP4B1* was not associated with IS, the role of hydroxyeicosatetraenoic acids and other CYP450-derived eicosanoids in brain vascular function and stroke has been extensively studied.³⁴ Moreover, genetic polymorphisms in the adjacent *CYP4A11* and other key enzymes of the CYP450 pathway have been associated with IS.^{35–37}

Our study has several strengths, including a well-characterized diverse cohort with detailed clinical and exposure data over a long period of follow-up. The prospective design allowed us to measure serum metabolite levels years or even decades before the clinical event; its treatment and comorbidities may have confounded analyses. In addition, we extended the findings to the acute stroke setting. However, we acknowledge some limitations. First, only strokes severe enough to warrant hospitalization are ascertained in ARIC. Second, sample size limitations on the number of subtype-specific strokes, most notably LAA, does not allow us to convincingly exclude the possibility that the observed associations also extend to LAA. Third, metabolomics profiling was performed at a single time point on biospecimens stored for an extended period of time. No data on metabolites' long-term stability were available. Variation in metabolite levels due to long-term storage conditions time can bias estimates of association between biomarkers and disease outcomes.³⁸ Thus, assuming that the degradation was nondifferential by stroke status, such changes may have resulted in a decreased power to detect associations. Fourth, our sample included predominantly African American participants and had higher rates of hypertension and diabetes than the rest of the cohort. Fifth, our genetic analyses were limited to a small number of protein coding variants. A more comprehensive investigation of functional genetic variants influencing intracellular hexadecanedioate levels may prove informative. Finally, while the association of

hexadecanedioate with IS was further supported by data from the case-control CIRCULAS study, formal external validation will require additional replication from an independent population-based cohort study.

This study highlights the potential of metabolomics for discovering novel circulating biomarkers for stroke and identifying novel susceptibility genes for IS and its subtypes. Further studies are needed to shed light on the possible pathways and mechanisms underlying these associations and to further investigate their potential clinical applications.

Author contributions

D. Sun: analysis and interpretation of data, critical revision of manuscript for intellectual content. S. Tiedt: acquisition of data, analysis and interpretation of data, critical revision of manuscript for intellectual content. B. Yu: analysis and interpretation of data, critical revision of manuscript for intellectual content. R.F. Gottesman: acquisition of data, critical revision of manuscript for intellectual content. T.H. Mosley: acquisition of data, critical revision of manuscript for intellectual content. E. Boerwinkle: study concept and design, acquisition of data. M. Dichgans: study concept and design, acquisition of data, critical revision of manuscript for intellectual content. M. Fornage: study concept and design, acquisition of data, analysis and interpretation of data, drafted manuscript, study supervision.

Acknowledgment

The authors thank the staff and participants of the ARIC study. The authors also wish to acknowledge the contribution of the MEGASTROKE Consortium. A complete list of MEGASTROKE investigators and contributors is provided in the data available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.55hr85m>.

Study funding

The Atherosclerosis Risk in Communities (ARIC) Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute (NHLBI) contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). The MEGASTROKE project received funding from sources specified at <http://www.megastroke.org/acknowledgments.html>. Study funded by NIH grant R01-NS087541. Funding support for “Building on GWAS for NHLBI diseases: The US CHARGE Consortium” was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419). Metabolomics measurements were sponsored by the National Human Genome Research Institute (3U01HG004402-02S1). This project also received funding from the Vascular Dementia Research Foundation, the European Union’s Horizon 2020 research and innovation programme under grant agreements 666881, SVDs@target and 667375, CoSTREAM, the DFG (SyNergy and CRC 1123 (B3)) and the Josef Hackl Foundation.

Disclosure

The authors report no disclosures relevant to the manuscript. Go to Neurology.org/N for full disclosures.

Publication history

Received by *Neurology* October 20, 2017. Accepted in final form December 19, 2018.

References

- McKay J, Mensah GA. The Atlas of Heart Disease and Stroke. Geneva: World Health Organization; 2004.
- Jauch EC, Saver JL, Adams HP Jr, et al. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 2013;44:870–947.
- Latchaw RE, Alberts MJ, Lev MH, et al. Recommendations for imaging of acute ischemic stroke: a scientific statement from the American Heart Association. *Stroke* 2009;40:3646–3678.
- Jiang Z, Sun J, Liang Q, et al. A metabolomic approach applied to predict patients with cerebral infarction. *Talanta* 2011;84:298–304.
- Jung JY, Lee HS, Kang DG, et al. 1H-NMR-based metabolomics study of cerebral infarction. *Stroke* 2011;42:1282–1288.
- ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) study: design and objectives. *Am J Epidemiol* 1989;129:687–702.
- Jones SA, Gottesman RF, Shahar E, Wruck L, Rosamond WD. Validity of hospital discharge diagnosis codes for stroke: the Atherosclerosis Risk in Communities Study. *Stroke* 2014;45:3219–3225.
- Rosamond WD, Folsom AR, Chambless LE, et al. Stroke incidence and survival among middle-aged adults: 9-year follow-up of the Atherosclerosis Risk in Communities (ARIC) cohort. *Stroke* 1999;30:736–743.
- The National Survey of Stroke. National Institute of Neurological and Communicative Disorders and Stroke. *Stroke* 1981;12:11–191.
- Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009;150:604–612.
- Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. Integrated, non-targeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal Chem* 2009;81:6656–6667.
- Ohta T, Masutomi N, Tsutsui N, et al. Untargeted metabolomic profiling as an evaluative tool of fenofibrate-induced toxicology in Fischer 344 male rats. *Toxicol Pathol* 2009;37:521–535.
- Sankoh AJ, Huque MF, Dubey SD. Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Stat Med* 1997;16:2529–2542.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B* 1995;57:289–300.
- Tiedt S, Prestel M, Malik R, et al. RNA-Seq identifies circulating miR-125a-5p, miR-125b-5p, and miR-143-3p as potential biomarkers for acute ischemic stroke. *Circ Res* 2017;121:970–980.
- Sacco RL, Kasner SE, Broderick JP, et al. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 2013;44:2064–2089.
- R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Core Team; 2017.
- Shin SY, Fauman EB, Petersen AK, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet* 2014;46:543–550.
- Malik R, Chauhan G, Traylor M, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* 2018;50:524–537.
- Menni C, Metrustry SJ, Ehret G, et al. Molecular pathways associated with blood pressure and hexadecanedioate levels. *PLoS One* 2017;12:e0175479.
- Kamel H, Okin PM, Elkind MS, Iadecola C. Atrial fibrillation and mechanisms of stroke: time for a new model. *Stroke* 2016;47:895–900.
- Alonso A, Yu B, Qureshi WT, et al. Metabolomics and incidence of atrial fibrillation in African Americans: the Atherosclerosis Risk in Communities (ARIC) study. *PLoS One* 2015;10:e0142610.
- Wanders RJ, Komen J, Kemp S. Fatty acid omega-oxidation as a rescue pathway for fatty acid oxidation disorders in humans. *FEBS J* 2011;278:182–194.
- Vorkas PA, Shalhoub J, Lewis MR, et al. Metabolic phenotypes of carotid atherosclerotic plaques relate to stroke risk: an exploratory study. *Eur J Vasc Endovasc Surg* 2016;52:5–10.
- Alexander JJ, Snyder A, Tongstad JH. Omega-oxidation of monocarboxylic acids in rat brain. *Neurochem Res* 1998;23:227–233.
- Menni C, Graham D, Kastenmuller G, et al. Metabolomic identification of a novel pathway of blood pressure regulation involving hexadecanedioate. *Hypertension* 2015;66:422–429.
- Bar-Tana J, Ben-Shoshan S, Blum J, et al. Synthesis and hypolipidemic and anti-diabetogenic activities of beta,beta,beta',beta'-tetrasubstituted, long-chain diolic acids. *J Med Chem* 1989;32:2072–2084.

28. Hermesh O, Kalderon B, Bar-Tana J. Mitochondria uncoupling by a long chain fatty acyl analogue. *J Biol Chem* 1998;273:3937–3942.
29. Boehme AK, Esenwa C, Elkind MS. Stroke risk factors, genetics, and prevention. *Circ Res* 2017;120:472–495.
30. Finsterer J, Kothari S. Cardiac manifestations of primary mitochondrial disorders. *Int J Cardiol* 2014;177:754–763.
31. Wolke C, Bukowska A, Goette A, Lendeckel U. Redox control of cardiac remodeling in atrial fibrillation. *Biochim Biophys Acta* 2015;1850:1555–1565.
32. Yee SW, Giacomini MM, Hsueh CH, et al. Metabolomic and genome-wide association studies reveal potential endogenous biomarkers for OATP1B1. *Clin Pharmacol Ther* 2016;100:524–536.
33. Kroetz DL, Xu F. Regulation and inhibition of arachidonic acid omega-hydroxylases and 20-HETE formation. *Annu Rev Pharmacol Toxicol* 2005;45:413–438.
34. Huang H, Al-Shabrawey M, Wang MH. Cyclooxygenase- and cytochrome P450-derived eicosanoids in stroke. *Prostaglandins Other Lipid Mediat* 2016;122:45–53.
35. Ding H, Cui G, Zhang L, et al. Association of common variants of CYP4A11 and CYP4F2 with stroke in the Han Chinese population. *Pharmacogenet Genomics* 2010;20:187–194.
36. Fornage M, Lee CR, Doris PA, et al. The soluble epoxide hydrolase gene harbors sequence variation associated with susceptibility to and protection from incident ischemic stroke. *Hum Mol Genet* 2005;14:2829–2837.
37. Yi X, Liao D, Fu X, Zhang B, Wang C. Interaction among CYP2C8, EPHX2, and CYP4A11 gene variants significantly increases the risk for ischemic stroke in Chinese populations. *J Atheroscler Thromb* 2015;22:1148–1157.
38. Kugler KG, Hackl WO, Mueller LA, Fiegl H, Graber A, Pfeiffer RM. The impact of sample storage time on estimates of association in biomarker discovery studies. *J Clin Bioinformatics* 2011;1:9–17.