

## Machine learning reveals serum sphingolipids as cholesterol-independent biomarkers of coronary artery disease

Annelise M. Poss, ... , Scott A. Summers, Mary C. Playdon

*J Clin Invest.* 2020;130(3):1363-1376. <https://doi.org/10.1172/JCI131838>.

Clinical Research and Public Health

Metabolism

Vascular biology

Ceramides are sphingolipids that play causative roles in diabetes and heart disease, with their serum levels measured clinically as biomarkers of cardiovascular disease (CVD).

We performed targeted lipidomics on serum samples from individuals with familial coronary artery disease (CAD) ( $n = 462$ ) and population-based controls ( $n = 212$ ) to explore the relationship between serum sphingolipids and CAD, using unbiased machine learning to identify sphingolipid species positively associated with CAD.

Nearly every sphingolipid measured ( $n = 30$  of 32) was significantly elevated in subjects with CAD compared with measurements in population controls. We generated a novel sphingolipid-inclusive CAD risk score, termed SIC, that demarcates patients with CAD independently and more effectively than conventional clinical CVD biomarkers including serum LDL cholesterol and triglycerides. This new metric comprises several minor lipids that likely serve as measures of flux through the ceramide biosynthesis pathway rather than the abundant deleterious ceramide species that are included in other ceramide-based scores.

This study validates serum ceramides as candidate biomarkers of CVD and suggests that comprehensive sphingolipid panels should be considered as measures of CVD.

The NIH (DK112826, DK108833, DK115824, DK116888, and DK116450); the Juvenile Diabetes Research Foundation (JDRF 3-SRA-2019-768-A-B); the American Diabetes Association; the American Heart Association; the Margolis Foundation; the National Cancer Institute, NIH (5R00CA218694-03); and the Huntsman Cancer Institute Cancer Center Support Grant (P30CA040214).

**Find the latest version:**

<https://jci.me/131838/pdf>



# Machine learning reveals serum sphingolipids as cholesterol-independent biomarkers of coronary artery disease

Annelise M. Poss,<sup>1,2</sup> J. Alan Maschek,<sup>3,4,5</sup> James E. Cox,<sup>3,4,5</sup> Benedikt J. Hauner,<sup>6,7</sup> Paul N. Hopkins,<sup>8</sup> Steven C. Hunt,<sup>8,9</sup> William L. Holland,<sup>1,2</sup> Scott A. Summers,<sup>1,2</sup> and Mary C. Playdon<sup>1,2,6</sup>

<sup>1</sup>Department of Nutrition and Integrative Physiology and <sup>2</sup>Diabetes and Metabolism Research Center, University of Utah School of Medicine, Salt Lake City, Utah, USA. <sup>3</sup>Department of Biochemistry, <sup>4</sup>Metabolomics Core Research Facility, <sup>5</sup>Proteomics Core Research Facility, and <sup>6</sup>Division of Cancer Population Sciences, Huntsman Cancer Institute, Salt Lake City, Utah, USA. <sup>7</sup>Department of Mathematics, Technical University of Munich, Munich, Germany. <sup>8</sup>Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah, USA. <sup>9</sup>Department of Genetic Medicine, Weill Cornell Medicine, Doha, Qatar.

**BACKGROUND.** Ceramides are sphingolipids that play causative roles in diabetes and heart disease, with their serum levels measured clinically as biomarkers of cardiovascular disease (CVD).

**METHODS.** We performed targeted lipidomics on serum samples from individuals with familial coronary artery disease (CAD) ( $n = 462$ ) and population-based controls ( $n = 212$ ) to explore the relationship between serum sphingolipids and CAD, using unbiased machine learning to identify sphingolipid species positively associated with CAD.

**RESULTS.** Nearly every sphingolipid measured ( $n = 30$  of 32) was significantly elevated in subjects with CAD compared with measurements in population controls. We generated a novel sphingolipid-inclusive CAD risk score, termed SIC, that demarcates patients with CAD independently and more effectively than conventional clinical CVD biomarkers including serum LDL cholesterol and triglycerides. This new metric comprises several minor lipids that likely serve as measures of flux through the ceramide biosynthesis pathway rather than the abundant deleterious ceramide species that are included in other ceramide-based scores.

**CONCLUSION.** This study validates serum ceramides as candidate biomarkers of CVD and suggests that comprehensive sphingolipid panels should be considered as measures of CVD.

**FUNDING.** The NIH (DK112826, DK108833, DK115824, DK116888, and DK116450); the Juvenile Diabetes Research Foundation (JDRF 3-SRA-2019-768-A-B); the American Diabetes Association; the American Heart Association; the Margolis Foundation; the National Cancer Institute, NIH (5R00CA218694-03); and the Huntsman Cancer Institute Cancer Center Support Grant (P30CA040214).

## Introduction

Coronary artery disease (CAD) is the most common type of cardiovascular disease (CVD) worldwide and the leading cause of death in the Western Hemisphere (1). The condition gives rise to atherosclerosis and ischemia, which contribute to arrhythmia, myocardial infarction (MI), heart failure, and sudden death (2). A family history of CAD is an independent risk factor for MI, and once a MI occurs, the patient is at greatly increased risk for subsequent adverse cardiac events. In addition to incurring a sub-

stantial individual health burden, CVD is the costliest disease in the United States, producing an economic toll that is projected to grow substantially over the coming decades (3). The combination of personal and financial costs necessitates improved means for identifying at-risk individuals in order to enhance patient care and optimize resource management.

CAD is multifactorial by nature, with obesity, diet, hypertension, type 2 diabetes mellitus, and family history of CVD established as risk factors (3). Traditional serum lipid biomarkers of CV health include triglycerides (TGs) and cholesterol, 2 abundant and easily quantifiable circulating factors. Recent technological advances now allow for the detection of less plentiful lipids, such as sphingolipids, enabling substantially more diverse lipidomic screenings at relatively high throughput. In leveraging these technological developments, researchers have identified a small subset of serum ceramides as biomarkers of CVD risk (4). Moreover, a substantial body of literature in rodent models of CVD indicates that these sphingolipids play causative roles in diabetes and cardiometabolic disorders (5).

### ► Related Commentary: p. 1084

**Authorship note:** SAS and MCP contributed equally to this work.

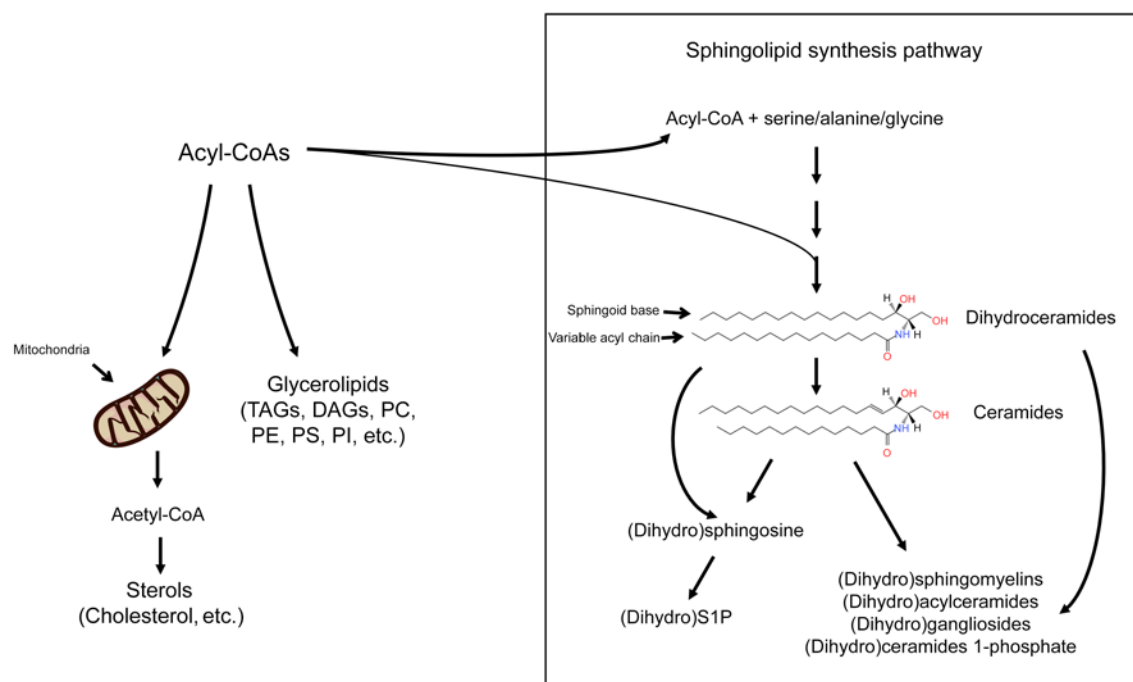
**Conflict of interest:** SAS is a cofounder of, consultant to, and shareholder in Centaurus Therapeutics.

**Copyright:** © 2020, American Society for Clinical Investigation.

**Submitted:** July 17, 2019; **Accepted:** November 13, 2019; **Published:** January 27, 2020.

**Reference information:** *J Clin Invest.* 2020;130(3):1363–1376.

<https://doi.org/10.1172/JCI131838>.



**Figure 1. Schematic depicting the sphingolipid biosynthesis pathway.** Fatty acyl-CoAs have 3 primary fates: entering the mitochondria to be used for energy via oxidation, to form glycerolipids for use in storage or membrane formation, or to be coupled to an amino acid and enter the sphingolipid biosynthesis pathway. Sphingolipids are a diverse class of lipids that represent a minor subset of the lipidome but play critical roles in signaling events. DAGs, diacylglycerols; PC, phosphocholine; PE, phosphoethanolamine; PI, phosphatidylinositol; PS, phosphatidylserine; TAGs, triacylglycerols.

Sphingolipids constitute a class of lipids that have diverse structural and signaling functions and discrete biological roles and tissue distributions. Their excessive accumulation occurs when the delivery of fatty acids exceeds the storage capacity or energy needs of a cell (Figure 1 and ref. 5), with the primary steps of *de novo* synthesis occurring in the ER (6). Tissue inflammation further increases ceramide biosynthesis rates (7). In the third step of the sphingolipid biosynthesis pathway, a family of (dihydro)ceramide synthases add variable acyl chains to a sphingoid scaffold to produce the dihydroceramides and, subsequently, ceramides, which are the key foundational unit of predominant sphingolipids (8). The dihydroceramides and ceramides can be further modified in the Golgi apparatus by the addition of various head groups, generating complex sphingolipids such as sphingomyelins and glucosylceramides. Ceramides, but not dihydroceramides, containing either C16:0 or C18:0 acyl chains drive insulin resistance and hepatic steatosis (7, 9–14). Other deleterious effects of ceramides that are relevant to CVD include the retention of lipoproteins in the vascular wall, impaired vasodilation, and induction of cardiomyocyte apoptosis (15).

Prior profiling studies have identified 3 ceramide species [i.e., cer(d18:1/16:0), cer(d18:1/18:0), and cer(d18:1/24:0)] that are positively associated with CVD incidence (16), secondary CVD events (17), and mortality (18–20). Cer(d18:1/24:0) has been reported to be negatively associated with CV death (18–20), but its relationship with CVD incidence is less clear. We reasoned that less abundant sphingolipids may serve as strong markers of flux through the biosynthesis pathway. Therefore, we performed an inclusive sphingolipid screen (32 sphingolipids) in individuals with CAD and in

population-based control subjects (Table 1). By applying variable selection techniques, we used these data to develop a superior sphingolipid-based score that demarcated individuals with CAD.

## Results

**Individual ceramides and CAD.** We quantified 32 sphingolipids including the major ceramides [cer(d18:1)], dihydroceramides [dihydro-cer(d18:0)], glucosylceramides [(glucosyl-cer(d18:1)], dihydrosphingomyelins [dihydro-SM(d18:0)], sphingomyelins [SM(d18:1)], sphinganine, and sphingosine (Figure 2). All sphingolipids measured, except for 2 glucosylceramides, were elevated in patients with CAD compared with levels in control subjects (Table 2). Sphingosine ( $P < 2 \times 10^{-16}$ ), dihydro-cer(d18:0/16:0) ( $P < 2 \times 10^{-16}$ ), dihydro-cer(d18:0/18:0) ( $P < 2 \times 10^{-16}$ ), and cer(d18:1/24:1) ( $P < 2 \times 10^{-16}$ ) were most strongly associated with CAD (OR<sub>per SD</sub> 3.47, 95% CI: 2.63–4.69; OR<sub>per SD</sub> 2.54, 95% CI: 2.06–3.18; OR<sub>per SD</sub> 2.82, 95% CI: 2.24–3.60; OR<sub>per SD</sub> 2.30, 95% CI: 2.24–3.60; OR<sub>per SD</sub> 2.29, 95% CI: 1.86, 2.85, respectively). Figure 3 depicts the ORs for CAD for all sphingolipid species measured, including the unadjusted model, a parsimonious model (i.e., a minimally adjusted model that includes the covariates age, sex, and BMI), and a fully adjusted model (i.e., a model that includes the covariates age, sex, BMI, total cholesterol [total-C], LDL cholesterol [LDL-C], HDL cholesterol [HDL-C], VLDL cholesterol [VLDL-C], TGs, hypertension, diabetes, and smoking).

**Ceramide risk score and CAD.** For each subject, we calculated the ceramide risk score (i.e., cardiac event risk test 1 [CERT1]) that was developed by Zora Biosciences and is in operation at the Mayo Clinic as a means of predicting 5-year risk of CV mortality (4, 21,

**Table 1. Baseline characteristics of patient and control participants in the Utah CAD study**

	Controls	Patients	P value
No. of subjects	212	462	
Sex			
Male, n (%)	91 (43%)	356 (77%)	
Age (yr) <sup>A</sup>	53.5 ± 6.9	55.6 ± 7.5	0.004
BMI <sup>A</sup>	28.3 ± 5.7	29.1 ± 5.2	0.040
NA, n (%)		16 (3.5%)	
Smoking			< 0.001
Yes, n (%)	43 (20%)	208 (45%)	
No, n (%)	169 (80%)	254 (55%)	
Diabetes			< 0.001
Yes, n (%)	11 (5%)	108 (23%)	
No, n (%)	201 (95%)	354 (77%)	
Hypertension			< 0.001
Yes, n (%)	54 (26%)	262 (57%)	
No, n (%)	158 (74%)	200 (43%)	
Lipid-lowering medication			0.187
Yes, n (%)	13 (6%)	44 (10%)	
No, n (%)	199 (94%)	418 (90%)	
Total-C (mg/dL) <sup>A</sup>	189.5 ± 3.3	209.2 ± 4.6	< 0.001
HDL (mg/dL)	46.7 ± 12.8	40.9 ± 1.2	< 0.001
LDL (mg/dL)	103.1 ± 2.8	128.7 ± 4.2	< 0.001
VLDL (mg/dL)	37.3 ± 22.5	39.3 ± 3	0.326
Serum TGs (mg/dL) <sup>A</sup>	178 ± 9.5	202.7 ± 1.4	0.008

Clinical characteristics of Utah CAD study patients ( $n = 212$ ) and controls ( $n = 462$ ). Variables were compared between patients and controls using a 2-tailed  $t$  test for continuous variables and a  $\chi^2$  test for categorical variables, with a  $P$  value of less than 0.05 considered significant. <sup>A</sup>Mean ± SD. NA, data not available.

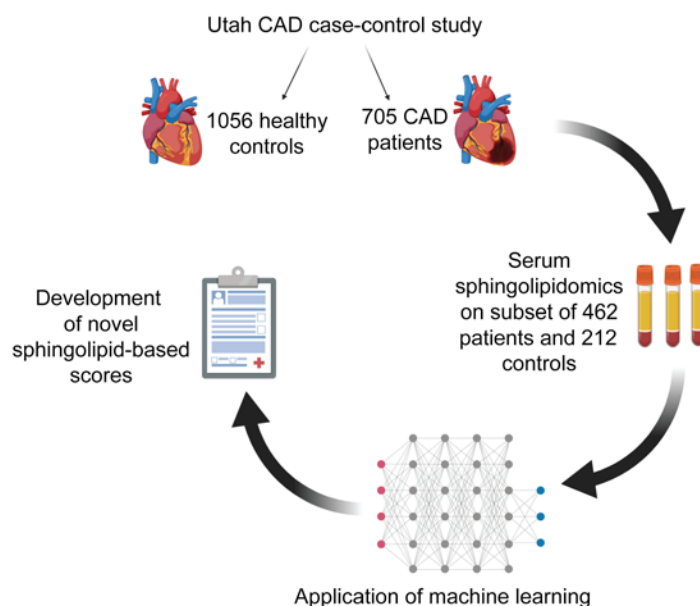
22). CERT1 performed well in this cohort, as subjects with CAD had significantly higher CERT1 risk scores than did the control participants ( $OR_{per\ SD} 2.18$ , 95% CI 1.77–2.71) (Figure 3). Interestingly, the CERT1 score, which comprises the individual ceramide species cer(d18:1/16:0), cer(d18:1/18:0), and cer(d18:1/24:1) as well as the ratio of these lipids to cer(d18:1/24:0), did not provide better predictive power than the individual ceramide species included in the score [cer(d18:1/16:0);  $OR_{per\ SD} 2.30$ , 95% CI: 1.87–2.6; cer(d18:1/18:0);  $OR_{per\ SD} 2.30$ , 95% CI: 1.87–2.85; cer(d18:1/24:1);  $OR_{per\ SD} 2.29$ , 95% CI: 1.86–2.85] (Figure 3). Since cer(d18:1/24:0) was also elevated in individuals with CAD ( $OR_{per\ SD} 2.12$ , 95% CI: 1.73–2.61), its inclusion in the denominator of CERT1 diminished the score's predictive power in our sample (Figure 4).

**Probing the role of specific ceramide species in CAD.** To discern how the chemical composition of sphingolipids influenced their association with CAD, we grouped them into 2 different categories. In 1 category, we summed all species within

a sphingolipid class (e.g., ceramides, dihydroceramides, sphingomyelins, etc.), independent of acyl chain length. In a second category, we summed all sphingolipids that had certain acyl chains attached to the sphingoid base (e.g., all species with C16:0, C18:0, C20:0, C24:1:0, or C24:0 acyl chains), independent of sphingolipid class. We found that total C24:1-containing sphingolipids ( $OR_{per\ SD} 2.66$ , 95% CI: 2.12–3.38) and/or total dihydroceramides, independent of chain length ( $OR_{per\ SD} 2.46$ , 95% CI: 1.99–3.10), were most strongly associated with CAD (Figure 5).

**Ceramide correlations with cholesterol and other conventional biomarkers.** In order to explore the relationship between ceramides and other common biomarkers of CVD risk, we generated a Gaussian graphical model (GGM) between ceramides, TGs, LDL-C, HDL-C, and VLDL-C (Figure 6). The GGM measured the correlation of sphingolipids with each other and with traditional lipid biomarkers. All correlations were conditioned on the presence of the other analytes ( $r \geq 0.20$ ), thus representing direct relationships that are uninfluenced by other components. The GGM demonstrated that ceramide species correlated with each other in a single, interconnected network but that their associations with classic CVD risk biomarkers were weak (i.e.,  $r < 0.20$ ). In Figure 6, the strength of the correlations is depicted by the thickness of the lines connecting lipid nodes. The strongest positive correlations (red lines) were between cer(d18:1/20:0) and cer(d18:1/18:0); dihydro-cer(d18:0/24:0) and dihydro-cer(d18:0/22:0); and dihydro-SM(d18:0/24:0) and dihydro-SM(d18:0.22.0) (Figure 6). As expected, VLDL-C positively correlated with TGs (23). Ceramides did not correlate with VLDL-C, TGs, or other lipid markers of CVD risk. These findings indicate that sphingolipids are largely independent of traditional CVD lipid biomarkers and therefore provide new information about disease status, a critical consideration when developing novel biomarkers.

**Generating novel CAD predictive ceramide risk scores using machine learning.** We used machine learning, a branch of artificial intelligence, to reduce our large set of sphingolipids to a small set of predictive biomarkers. Machine learning incorporates pattern



**Figure 2. Schematic of the Utah CAD study design and the subset of available biospecimens used for LC-MS/MS sphingolipid analysis.** Machine learning was applied to the sphingolipidomic data to develop novel scores that associated with CAD beyond conventional lipid markers, such as cholesterol (created with BioRender).

**Table 2. Means and interquartile ranges for LC-MS/MS measured sphingolipids in and control groups of the Utah CAD study**

Lipid	Controls	Patients	P value
Dihydro-cer d18:0,16:0	0.1 (0.8–0.2)	0.2 (0.2–0.3)	$<2 \times 10^{-16}$
Dihydro-cer d18:0,18:0	0.09 (0.06–0.1)	0.1 (0.09–0.2)	$<2 \times 10^{-16}$
Dihydro-cer d18:0,20:0	0.05 (0.03–0.07)	0.07 (0.05–0.1)	$2.36 \times 10^{-10}$
Dihydro-cer d18:0,22:0	0.2 (0.1–0.3)	0.3 (0.2–0.5)	$1.39 \times 10^{-12}$
Dihydro-cer d18:0,24:0	0.4 (0.3–0.7)	0.7 (0.4–1.2)	$1.29 \times 10^{-14}$
Dihydro-cer d18:0,24:1	0.2 (0.1–0.4)	0.4 (0.2–0.7)	$4.26 \times 10^{-15}$
Cer d18:1,16:0	131.3 (87.8–201.3)	217.2 (150.3–324.9)	$5.83 \times 10^{-16}$
Cer d18:1,18:0	48 (29.9–76.7)	86 (53.9–138.3)	$5.40 \times 10^{-16}$
Cer d18:1,20:0	44.1 (30.6–65.6)	68.1 (43.8–102.3)	$6.12 \times 10^{-13}$
Cer d18:1,22:0	264.4 (185.9–436.7)	399.1 (278.8–631.3)	$3.63 \times 10^{-11}$
Cer d18:1,24:0	98 (64.5–148.3)	157.6 (106.8–245.1)	$1.61 \times 10^{-15}$
Cer d18:1,24:1	264.7 (173.5–411.8)	437.6 (306.2–669.2)	$<2 \times 10^{-16}$
GC-cer d18:1,16:0	364.7 (293.8–466.3)	366.3 (295–454.3)	0.98
GC-cer d18:1,18:0	64.4 (45.4–89)	70.7 (52.8–94.7)	0.30
GC-cer d18:1,20:0	66.3 (45.1–95.6)	94 (65.1–138.2)	$3.34 \times 10^{-8}$
GC-cer d18:1,22:0	494.7 (367.7–705)	713.5 (446.5–1022.2)	$2.87 \times 10^{-8}$
GC-cer d18:1,24:0	456.9 (321.1–591.1)	585.5 (408.2–879.4)	$8.14 \times 10^{-8}$
GC-cer d18:1,24:1	397.5 (287.4–547.1)	575.7 (399.9–866.1)	$5.11 \times 10^{-10}$
Dihydro-SM d18:0,16:0	45 (30.9–64.1)	61.1 (45.9–88.9)	$1.22 \times 10^{-9}$
Dihydro-SM d18:0,18:0	14 (7.5–25.2)	26.37 (14.3–49.6)	$2.72 \times 10^{-9}$
Dihydro-SM d18:0,20:0	28.3 (12.9–48.2)	48.4 (25–91.2)	$1.25 \times 10^{-7}$
Dihydro-SM d18:0,22:0	4.8 (2.2–10.2)	10.2 (5.9–21.4)	$9.08 \times 10^{-10}$
Dihydro-SM d18:0,24:0	0.9 (0.4–1.5)	1.5 (0.9–2.8)	$1.57 \times 10^{-10}$
Dihydro-SM d18:0,24:1	21.7 (10.7–42.1)	47.6 (26.3–92.2)	$1.40 \times 10^{-10}$
SM d18:1,16:0	592 (437.7–818.5)	779.3 (595–1066.3)	$3.77 \times 10^{-10}$
SM d18:1,18:0	155.1 (108.3–217.6)	186.7 (136.4–268.6)	$2.54 \times 10^{-6}$
SM d18:1,20:0	183 (78.7–354.3)	354.7 (183.3–698.6)	$1.03 \times 10^{-7}$
SM d18:1,22:0	359.4 (150.9–740)	718.7 (365.1–1462.2)	$4.28 \times 10^{-9}$
SM d18:1,24:0	154.2 (70.4–283.4)	293.8 (174.2–640.8)	$1.44 \times 10^{-9}$
SM d18:1,24:1	432.8 (200.5–879.1)	913.1 (478.2–1907.4)	$1.85 \times 10^{-9}$
Sphinganine	0.03 (0.02–0.04)	0.04 (0.03–0.06)	$4.92 \times 10^{-6}$
Sphingosine	0.08 (0.05–0.1)	0.15 (0.1–0.3)	$<2 \times 10^{-16}$

A 2-tailed *t* test was used to compare concentrations of sphingolipids measured by LC-MS/MS for patients (*n* = 212) and controls (*n* = 462). *P* values are for the parsimonious age-, sex-, BMI-adjusted model and were considered significant at a FDR of less than 0.05. Lipid concentrations are represented here as the mean (IQR). The fully adjusted model (i.e., age, sex, BMI, total-C, LDL-C, VLDL-C, TGs, hypertension, diabetes, smoking) was also run, but the results were not materially different from those of the parsimonious model, so only data from the minimally adjusted model are presented here. Cer, ceramide; GC-cer, glucosylceramide. Units are pmol lipid/mL serum.

recognition within complex data sets and has been used previously to develop CVD risk prediction models. In comparison with classical statistical methods, machine-learning techniques can identify algorithms that predict health outcomes, even when relationships are complex and nonlinear (24, 25). Moreover, machine-learning-generated models tend to be more generalizable (24, 25).

We created these new sphingolipid-based risk scores using random forest (RF) and least absolute shrinkage and selection operator (LASSO) regression approaches for variable reduction and selection (Table 3). The RF method develops algorithms that can precisely classify observations into groups (i.e., CAD patients versus controls). With this method, the number of variables incorporated has a strong impact on model accuracy: if variables improve the

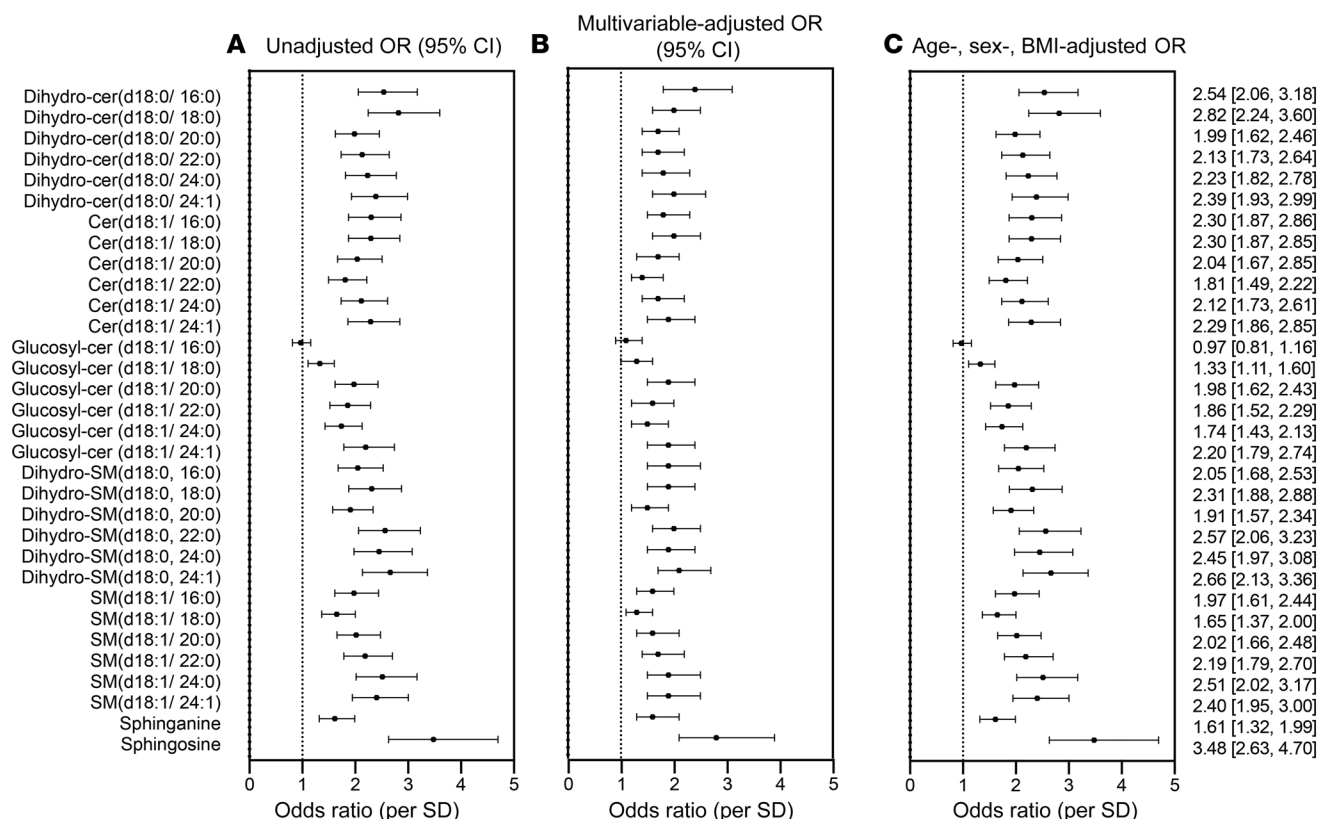
model fit, RF accuracy improves; if not, accuracy is diluted by meaningless variables. We therefore ran 2 RF models. For the first, our input included sphingolipid variables only. For the second, our input included sphingolipid variables in concert with classical CVD risk markers (LDL-C, HDL-C, VLDL-C, and TGs). For our LASSO approach, the input included all sphingolipids and the aforementioned conventional CVD lipid markers. We evaluated the biomarker score classification using both ORs and receiver operator characteristic-area under the curve (ROC-AUC) analysis (Figure 7 and Table 4). For both RF and LASSO approaches, the 5 lipids most positively associated with CAD were used to generate a score.

An RF-generated sphingolipid-inclusive CAD (RF-SIC) risk score (AUC = 0.75) outperformed CERT1 (AUC = 0.67) and conventional CVD risk biomarkers including LDL-C (AUC = 0.69) and total-C (AUC = 0.63) (Figure 7 and Supplemental Figure 1; supplemental material available online with this article; <https://doi.org/10.1172/JCI131838DS1>). An RF model generated from the sphingolipids plus CVD risk markers (denoted with a superscript plus sign, RF-SIC<sup>+</sup>, AUC = 0.78) included LDL-C and displayed a more precise classification of CAD patients versus controls, as compared with the RF-SIC score that excluded LDL-C (Figure 7). When evaluated by OR, RF-SIC<sup>+</sup> (OR 5.03, 95% CI: 3.69–7.07) outperformed the RF-SIC score (OR 3.49, 95% CI: 2.71–4.58) (Figure 7).

The LASSO-generated SIC (LASSO-SIC) performed similarly to the RF-generated score (AUC for LASSO-SIC = 0.74; OR 2.86, 95% CI: 2.67, 3.66). We conducted an exploratory analysis, adding a term that was the ratio of the lipid with the highest positive CAD association versus the lipid that had the most negative association. This resulted in a slight increase in predictability (LASSO-SIC2, AUC = 0.75; OR 3.06, 95% CI: 2.42, 3.94) (Figure 7, Table 4, and Supplemental Figure 2). Adding in another ratio (i.e., the second-highest, positively associated lipid versus the second-highest, negatively associated lipid variables) enhanced performance further (LASSO-SIC3, AUC = 0.77; OR 3.91, 95% CI: 2.98, 5.24) (Figure 7, Table 4, and Supplemental Figure 2).

On the basis of this information, we generated a final SIC score that included the highest-performing sphingolipid RF- and LASSO-generated components and yielded increased discriminatory ability (AUC = 0.79; OR 4.67, 95% CI: 3.47, 6.43) (Figure 7 and Table 4). Only sphingolipids, and not LDL-C, were included in the final SIC score, so that comparisons of an inclusive sphingolipid measurement with conventional CVD lipid markers could





**Figure 3. Forest plot of OR (95% CI) for CAD per SD of sphingolipid species in the Utah CAD study.** (A) Unadjusted OR. (B) Fully adjusted OR (age, sex, BMI, total-C, LDL-C, VLDL-C, TGs, hypertension, diabetes, smoking). (C) Minimally adjusted OR (age, sex, BMI) model. The numerically presented ORs (95% CI) represent the minimally adjusted age, sex, and BMI model.

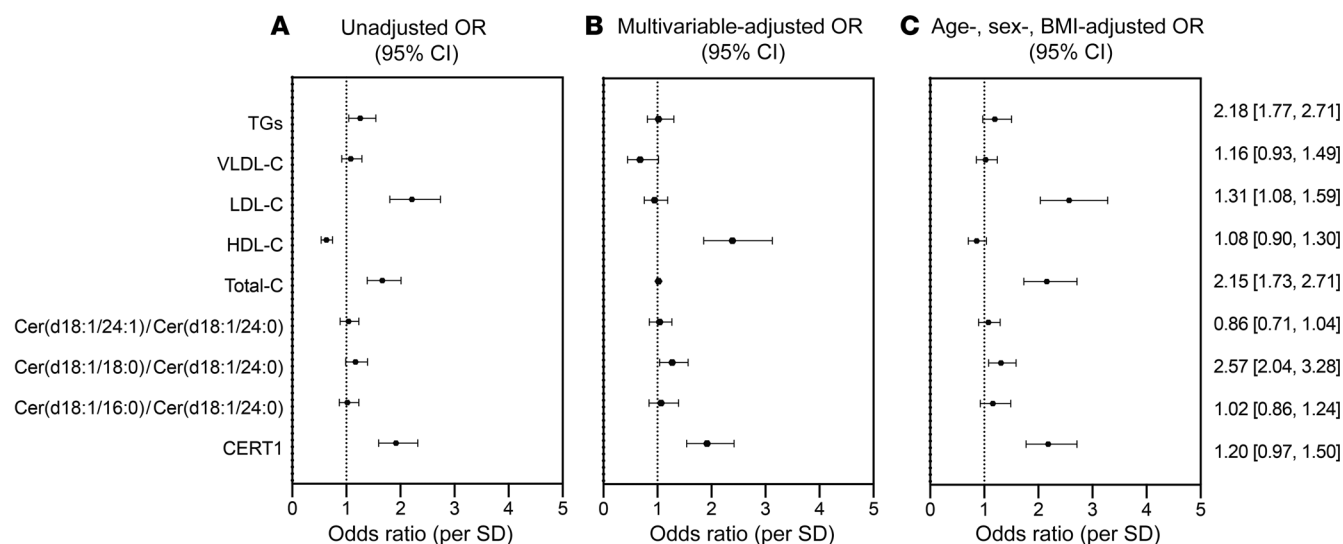
be performed. A list of the lipid components in each novel score is provided in Table 3.

**Comparison of machine-learning-generated scores with conventional markers of CAD.** We next compared the ability of SIC, CERT1, and standard clinical biomarkers (TGs, LDL-C, etc.) to classify CAD patients compared with controls (Figure 8A and Table 4). We provide the following ROC curves (with the AUC) for comparison (Figure 8, B–G): clinical factors alone (age, sex, BMI, diabetes, hypertension, smoking; AUC = 0.63); clinical factors plus CERT1 (AUC = 0.66); clinical factors plus SIC (AUC = 0.72); clinical factors plus standard clinical lipids (AUC = 0.64); CERT1 plus clinical factors and clinical lipids (AUC = 0.64); and SIC plus clinical factors and clinical lipids (AUC = 0.65). Since the AUC can be an insensitive measure of model performance, particularly when the initial model (i.e., American Heart Association/American College of Cardiology [AHA/ACC] risk factors) performs strongly, we also calculated a continuous net reclassification index (NRI) and an integrated discrimination index (IDI) (26). These scores provide a more comprehensive picture of model performance and a means to assess the value of including SIC or CERT1 in addition to standard clinical biomarkers. For SIC, the NRI was 0.67 (95% CI: 0.52–0.81,  $P < 0.0001$ ) and the IDI was 0.10 (95% CI: 0.08–0.11,  $P < 0.0001$ ) (Supplemental Table 1). As a frame of reference, an NRI exceeding 0.6 is considered strong and 0.4 is considered intermediate (27). The SIC was superior to CERT1, which had an NRI of 0.48 (95% CI: 0.32–0.64,  $P < 0.0001$ ) and an IDI of 0.04

(95% CI: 0.03–0.06,  $P < 0.0001$ ) (Supplemental Table 1). The SIC improved the ROC C-statistic, NRI, and IDI compared with AHA/ACC guideline risk factors alone, underscoring the power of including sphingolipids as biomarkers of CAD.

Many of the lipids extracted by our variable reduction techniques [i.e., SM(d18:0/24:1), SM(d18:0/22:0), SM(d18:0/18:0), sphingosine, cer(d18:0/18:0), and cer(d18:0/16:0)] are transient intermediate lipid species and therefore reflect pathway activity and flux (for a full list of selected lipids, see Table 3). This finding suggests that although abundant ceramide species are implicated in driving disease states, these causal lipid species may not be the most sensitive clinical markers.

**Stratification by CAD presentation.** To further probe the clinical utility of the SIC score, we evaluated it in patients with CAD who were stratified into 3 subgroups: (a) patients having had a MI only; (b) patients who had a surgical intervention only (coronary artery bypass grafting [CABG] or percutaneous transluminal coronary angioplasty [PTCA]); or, (c) patients who had an MI in combination with a surgical intervention. Patients undergoing a surgical intervention only are considered to have a more tightly controlled disease state, whereas those with both surgical intervention and MI are likely to be in a more severe or uncontrolled disease state (28). Patients with an MI only are considered intermediate. As compared with the control population (i.e., all non-cases), the CERT1 and SIC scores were highest in the individuals with the more severe disease presentation ( $OR_{\text{per SD}} > 1.80$ ,  $P < 5 \times 10^{-11}$ ;



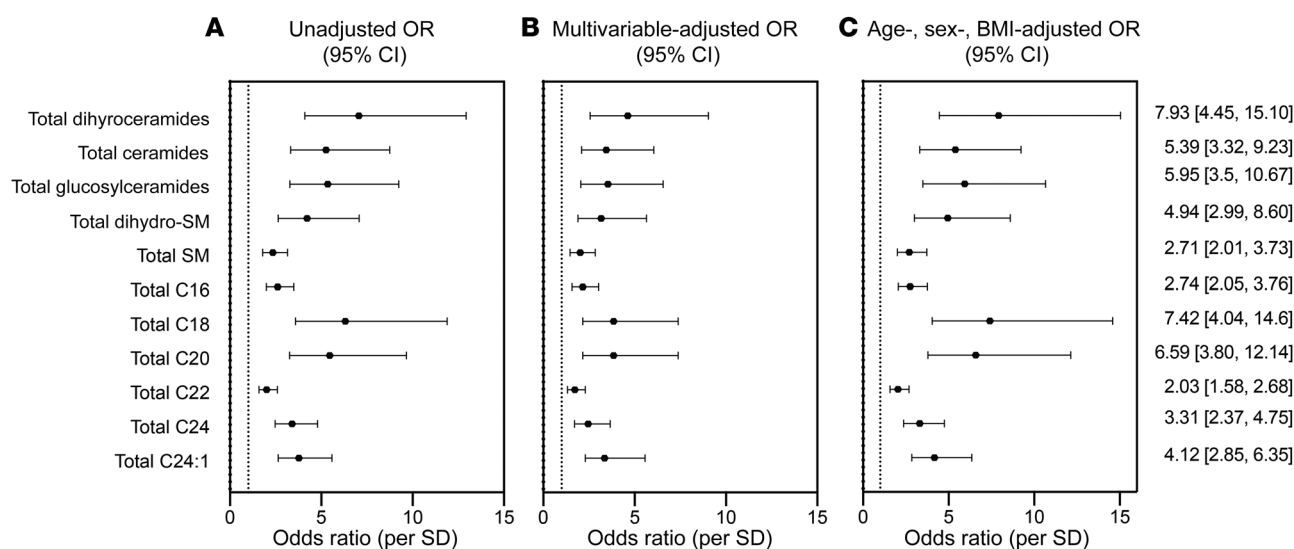
**Figure 4. OR (95% CI) of CAD per SD of previously reported lipid markers of CVD in the Utah CAD study.** (A) Unadjusted OR. (B) Fully adjusted OR (age, sex, BMI, hypertension, diabetes, smoking). (C) Minimally adjusted OR (age, sex, BMI). The numerically presented ORs (95% CI) represent the minimally adjusted age, sex, and BMI model. Since we compared clinical lipid markers (LDL, VLDL, HDL, TGs) with ceramide ratios and scores, they were not included in the fully adjusted model. CERT1, cardiac event risk test (12-point scale). HDL-C, LDL-C, VLDL-C, and TG values are given in mg/dL.

$P$  heterogeneity  $< 2 \times 10^{-16}$ ) (Figure 9). By comparison, standard clinical markers including LDL-C, total-C, and TGs did not show a preferential increase for individuals in this, as opposed to any other, category (Table 5). These findings suggest that ceramide-based scores may have utility for risk stratification, which is in line with previous studies that demonstrated the capacity of ceramides, but not LDL-C, to predict secondary cardiac events (17).

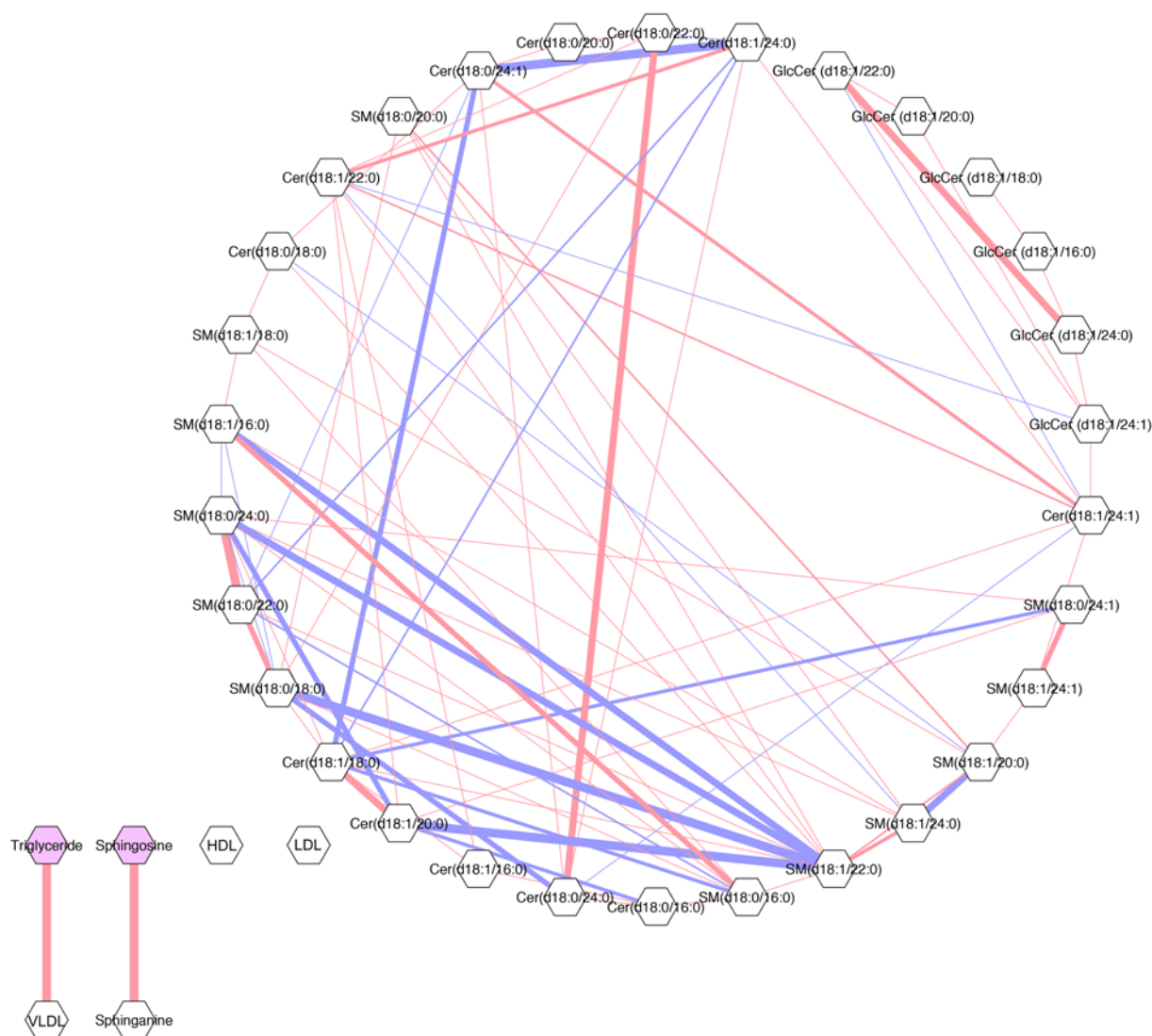
## Discussion

We applied a highly quantitative, targeted mass spectroscopy platform to measure 32 sphingolipids in serum samples from patients

with CAD compared with samples from healthy controls. Thirty of the 32 sphingolipids assayed were elevated among the diseased subjects, displaying a robust positive association with CAD after controlling for multiple comparisons. We applied unbiased machine-learning variable reduction techniques to generate a novel sphingolipid score we have termed the sphingolipid-inclusive CAD (SIC) risk score, which includes the following components: dihydro-cer(d18:0/18:0), cer(d18:1/18:0), cer(d18:1/22:0), cer(d18:1/24:0), dihydro-SM(d18:0/24:1), SM(d18:1/24:0), SM(d18:1/18:0), and sphingosine. Novel scores were calculated by summing raw lipid values multiplied by their  $\beta$  coefficients



**Figure 5. OR (95% CI) of CAD per SD of summed sphingolipid variables in the Utah CAD study.** (A) Unadjusted OR. (B) Fully adjusted OR (age, sex, BMI, total-C, LDL-C, VLDL-C, TGs, hypertension, diabetes, smoking). (C) Minimally adjusted OR (age, sex, BMI). The numerically presented ORs (95% CI) represent the minimally adjusted age, sex, and BMI model. Total SM, total sphingomyelin; Total C16, sum of all C16 acyl chains; Total C18, sum of all C18 acyl chains; Total C20, sum of all C20 acyl chains; Total C22, sum of all C22 acyl chains; Total C24, sum of all C24 acyl chains; Total C24:1, sum of all C24:1 acyl chains.



**Figure 6. GGM of correlations between ceramide species and conventional lipid markers in patients with CAD.** Conditioned on the presence of all other analytes ( $r \geq 0.20$ ). Analytes are represented by nodes (gray hexagons) and conditional correlations by edges (lines). Pink lines indicate positive correlations and blue lines indicate the inverse. Line width represents the strength of the conditional correlation, and the lack of a line indicates no detectable relationship above the threshold. GlcCer, glucosylceramide.

from the regression output and then log transformed. This score approached a strong C-statistic of 0.79 and an  $OR_{\text{per SD}}$  of 4.67 (95% CI: 3.46–6.43) for risk of CAD, outperforming other serum indices of CVD risk including LDL-C alone and the CERT1 ceramide risk score. Serum ceramides were also associated with disease severity, as they were highest among individuals with the most severe CAD manifestations. These findings support the idea that serum sphingolipids are strong biomarkers of CAD that could have clinical utility for improving risk stratification.

These data are consistent with several other studies using untargeted lipidomic platforms, which frequently identified sphingolipids as candidate biomarkers of CVD (17, 18, 20, 22, 29–31). Ceramide concentrations and scores were shown to be elevated among individuals with acute MI (16), CAD (22, 29, 32), acute coronary syndrome (22), and recurrent major adverse cardiac events (20). The concentrations were also increased in individuals with insulin resistance or type 2 diabetes (33–35), two underlying driv-

ers of CV morbidity (33, 34). These studies implicate ceramides as markers of disease pathology, disease risk, and mortality and as a tool for improved risk stratification. The best-characterized ceramide score is CERT1, originally developed by Zora Biosciences and validated in multiple prospective clinical studies (17, 18, 20, 22, 29–31). Though most of the ceramide species contained within CERT1 were individually predictive of CAD, they were not identified as the most strongly CAD-associated lipids using our unbiased variable selection methods. Furthermore, the CERT1 inclusion of cer(18:1/24:0) in the denominator was counterproductive; cer(18:1/24:0) was itself a good marker of CAD. Nonetheless, CERT1 still performed similarly in this data set as compared with previous prospective cohort studies, thus endorsing its validity as a robust index of CVD risk.

The most widely used biomarker of CV pathophysiology, LDL-C, also performed well in this data set. However, SIC and CERT1 showed stronger discriminatory power than did LDL-C as



**Table 3. Novel sphingolipid scores for CAD generated through the application of machine-learning techniques**

Score	Components
RF-SIC	Sphingosine, cer(d18:0/18:0), cer(d18:0/16:0), SM(d18:0/24:1), SM(d18:0/22:0)
RF-SIC*	LDL-C, sphingosine, cer(d18:0/18:0), cer(d18:0/16:0), SM(d18:1/24:0)
LASSO-SIC	SM(d18:0/24:1), cer(d18:1/18:0), cer(d18:1/24:0), cer(d18:1/18:0), SM(d18:1/24:0)
LASSO-SIC2	SM(d18:0/24:1) / SM(d18:1/18:0), cer(d18:1/18:0), cer(d18:1/24:0), cer(d18:1/18:0), SM(d18:1/24:0)
LASSO-SIC3	SM(d18:0/24:1) / SM(d18:1/18:0), cer(d18:1/18:0) / cer(d18:1/22:0), cer(d18:1/24:0), cer(d18:1/18:0), SM(d18:1/24:0)
SIC	SM(d18:0/24:1) / SM(d18:1/18:0), cer(d18:1/18:0) / cer(d18:1/22:0), cer(d18:1/24:0), cer(d18:1/18:0), SM(d18:1/24:0), sphingosine

RF and LASSO regression were applied for variable reduction and selection.

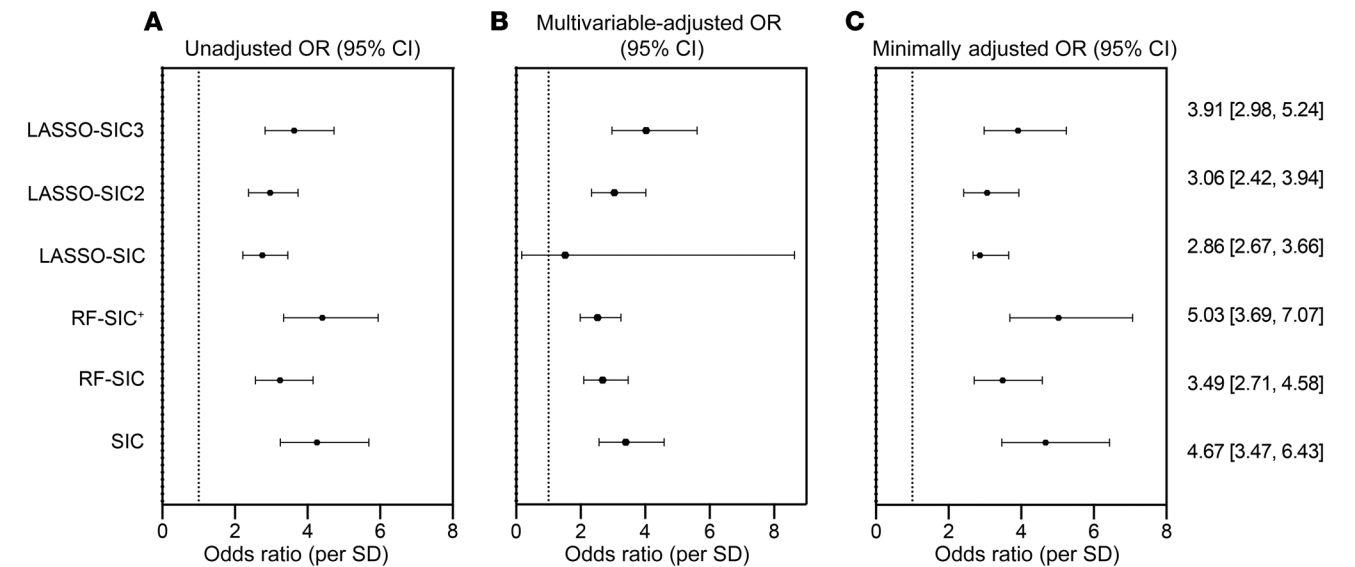
assessed by ROC-AUC. Interestingly, ceramides were not strongly correlated with LDL-C (<0.20), though LDL-C was strongly correlated with other conventional lipid markers such as serum TGs. The independence of these biomarkers is consistent with the idea that they lie in different biosynthetic pathways, both of which contribute to disease progression.

National screening and therapeutic guidelines focus on cholesterol as the primary biomarker of CV health, even though it shows only modest predictive utility for risk assessment and lacks the sensitivity to discriminate between patients at risk for secondary cardiac events (17). Current guidelines dictate that patients diagnosed with CAD belong to a high-risk population, even though this classification may be inaccurate for most individuals (17). By combining LDL-C with novel sphingolipid risk scores, a more complete risk assessment may be performed. Such a tool will enhance patient classification accuracy and help the clinician to coordinate disease surveillance or prescribe clinical interventions.

Ceramides are not only biomarkers of CV health but are probably causative agents in disease progression (15). Studies in rodent models revealed that pharmacological inhibition of ceramide synthesis prevents ischemic cardiomyopathy-related heart

failure, while simultaneously diminishing ventricular remodeling, fibrosis, and macrophage infiltration following MI (36–39). Moreover, such ceramide-lowering interventions resolve dyslipidemia, insulin resistance, hypertension, atherosclerosis, and hepatic steatosis (7, 40–49), conditions that underlie CVD. Manipulations of the de novo ceramide synthesis pathway further suggest that certain ceramide species are deleterious, whereas others are benign or beneficial (11–14); those containing the C16 or C18 acyl chain (11–13) and include the double bond (i.e., ceramides, not dihydroceramides) (7) in the sphingolipid backbone are particularly harmful. Last, studies in rodents revealed that ceramide degradation is a primary means by which adiponectin receptors, which are ligand-activated ceramidases (50), exert their antidiabetic, cardioprotective, and insulin-sensitizing actions (50–52). Cumulatively, these data identify ceramides as some of the more toxic metabolites accumulating in states of metabolic distress.

Our machine-learning variable reduction approaches (RF, LASSO) for score generation extracted sphingosine, dihydrocer(d18:0/16:0), dihydrocer(d18:0/18:1), dihydro-SM(d18:0/24:1), dihydro-SM(d18:0/22:0), SM(d18:1/18:0), cer(d18:1/18:0), and cer(d18:1/24:0) as the lipid species most positively associat-



**Figure 7. OR (95% CI) of CAD per SD of novel scores generated through the application of machine-learning approaches in the Utah CAD study. (A)** Unadjusted OR. **(B)** Multivariable-adjusted OR (age, sex, BMI, diabetes, hypertension, smoking). **(C)** Minimally adjusted OR (age, sex, BMI). The multivariable models for this analysis do not include HDL-C, LDL-C, VLDL-C, total-C, or TGs, as they were included as input variables.

**Table 4. AUC of ROC plots for lipid-based clinical indices**

Clinical index	C-statistic
TGs	0.54
LDL-C	0.69
CERT1	0.67
RF-SIC	0.75
RF-SIC*	0.78
LASSO-SIC	0.74
LASSO-SIC2	0.75
LASSO-SIC3	0.76
SIC	0.79

ROC curves were generated and C-statistics were calculated for each clinical index.

ed with CAD. This finding suggests that the more abundant ceramides, including those that have been established as drivers of tissue and metabolic dysfunction, may not be the most sensitive biomarkers for CAD. Rather, less-abundant lipids that serve as markers of increased ceramide biosynthetic flux may provide a more accurate and comprehensive readout of disease status.

Though some prior studies have described associations between a subset of ceramides and CVD and related comorbidities, we believe that several aspects of this study are novel. First, we conducted a comprehensive ceramide assessment using a well-validated, targeted lipidomic platform that included less-abundant lipid species, leading to the production of a more robust sphingolipid score (i.e., SIC). We note that such targeted platforms are more quantitatively sound than shotgun lipidomic assessments. Second, we focused on patients with early-onset CAD (average age of onset = 47.8 years), thus enhancing the power of our study and limiting the influence of factors associated with aging. Third, we applied machine learning to develop new ceramide-based scores that outperformed prior measures, including LDL-C and CERT1. Machine learning allowed us to enhance the accuracy of models and reduce the dimensionality of data sets (53).

Despite these advances, our study has some limitations. First, it is limited by its case-control design and by the racial homogeneity of our sample population, limiting generalizability. Second, our target lipid class, sphingolipids, includes highly diverse and lowly abundant lipid species; this diversity can lead to increased variability, as seen by our high coefficients of variation (median: 11.76, IQR: 6.85–20.53). Although these coefficients of variation are not ideal, they are comparable to those of previous sphingolipidomic studies. Third, this study lacks a validation cohort for the novel SIC score. We note, however, that this cohort recapitulated the findings relating to the CERT1 score, which was generated using alternative patient data sets. And fourth, some biospecimens were collected as far back as the 1990s; diet and lifestyle have changed since this study was initiated, and prolonged storage could negatively impact sample quality. Nonetheless, samples from patients and controls were collected and handled in the same manner, so relative differences (and calculated ORs for CAD) should be sustained. Moreover, prior studies have shown that sphingolipids remain stable over storage periods as long as 16 years after sample collection and through multiple freeze-thaw

cycles (54, 55). We emphasize the exciting fact that sphingolipids appear to serve as strong biomarkers across generations, as any robust clinical index should (56).

In conclusion, sphingolipids have emerged as robust, cholesterol-independent markers of CVD risk. Their inclusion in a clinician's armamentarium has the potential to greatly improve the ability to identify at-risk patients. Moreover, our findings support the development of therapeutics targeting sphingolipids as a means of ameliorating CVD. Nonetheless, our data suggest that further refinement of sphingolipid-based scores may be necessary. Expanding the diversity of sphingolipid entities included in prospective patient studies will provide a more complete picture of the sphingolipidome for the prediction CVD risk.

## Methods

**Study design.** We evaluated the association of serum sphingolipids with CAD using existing samples and clinical and demographic information obtained from a case-control study conducted in Utah, USA ( $n = 462$  patients;  $n = 212$  controls) (57).

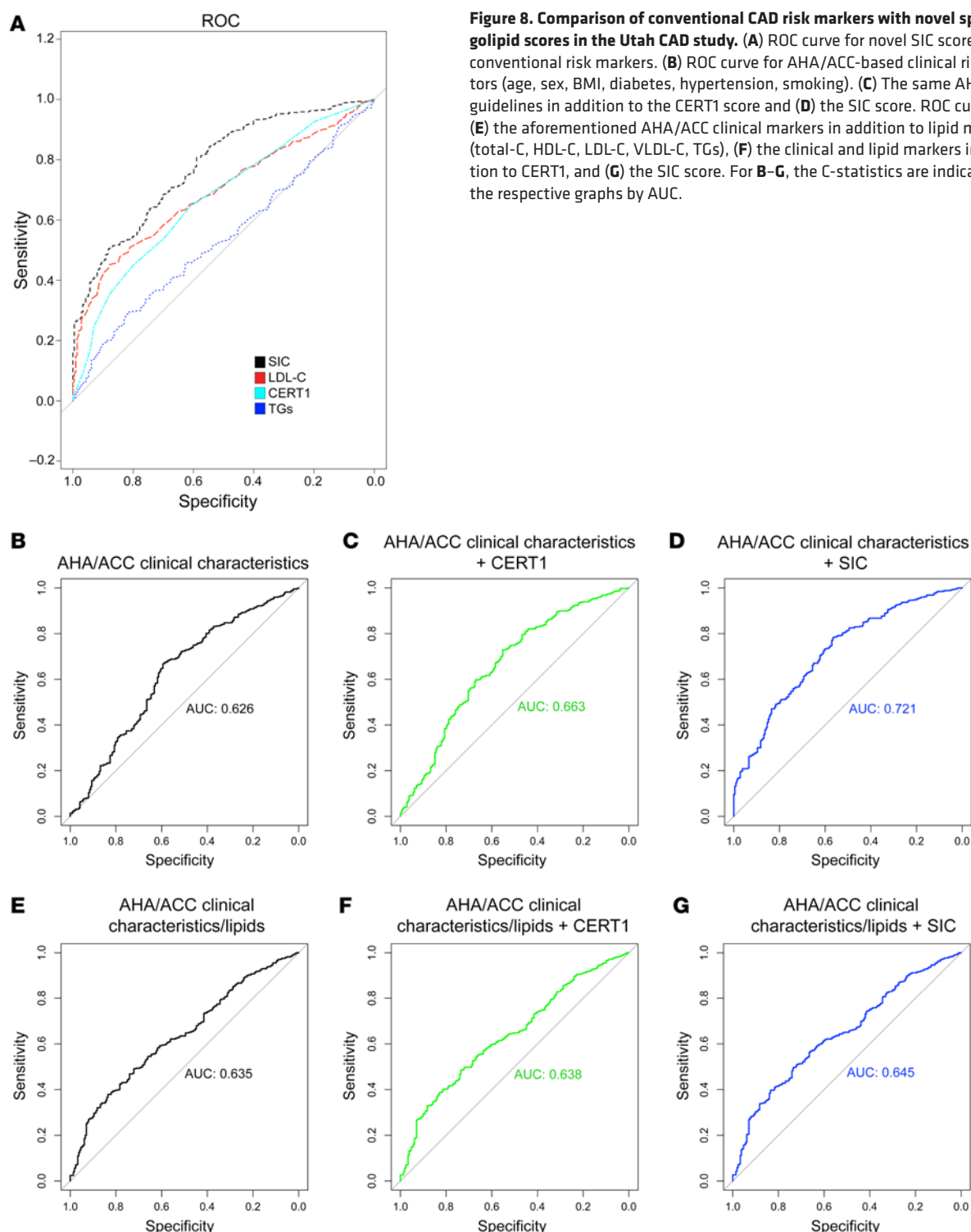
**Study population.** Patients were recruited between 1990 and 2000 from Intermountain Healthcare discharge records or the Family Health Tree Program in Utah (58). The patients were aged 30–75 years and had a diagnosis of CAD, defined by the original study recruitment criteria as MI, PTCA, or CABG. A large proportion of patients were male (77%), probably because premature CAD incidence rates are higher for men than women (59). The patients had an age of onset similar to that of at least 1 first-degree relative (parent, sibling, or child) (Table 1). To limit artifactual effects of the acute cardiac event on lipid levels, samples were collected at least 6 months following the event.

Control subjects representative of the Utah population (57, 60, 61) were randomly sampled from 1980–1986 from (a) the parents of students participating in the Family Health Tree Program (58), a study of family health among Utah high schools; and (b) spouse pairs participating in a study on psychological factors concerning CAD (61). The control participants were aged 30–75 years and had no clinical diagnosis of CAD, but they could have a family history of CAD. Control subjects taking vasoconstrictive drugs (i.e., beta blockers, calcium channel blockers, and other antianginal medications) were excluded.

Both patient and control populations were selected from the same source population of Salt Lake City, Utah. The number of patients ( $n = 462$ ) was larger than that of control subjects ( $n = 212$ ) because of the nature of the available biospecimens, though no significant differences between this subset of specimens available for analysis and the original study sample were noted (Supplemental Table 2).

**Clinical and demographic characteristics.** Demographic information (including age and sex) and medical and family history data were obtained by trained interviewers. Covariates considered in the analyses included age (years); sex (male or female); BMI ( $\text{kg}/\text{m}^2$ ), smoking (“ever” or “never” to smoking daily for 1 year or more); total-C ( $\text{mg}/\text{dL}$ ), LDL-C ( $\text{mg}/\text{dL}$ ), VLDL-C ( $\text{mg}/\text{dL}$ ), HDL-C ( $\text{mg}/\text{dL}$ ), TGs ( $\text{mg}/\text{dL}$ ); lipid medication (statins, fibrates, and other hyperlipidemia-managing drugs taken at the time of the blood draw, yes/no); diabetes (prior physician diagnosis or fasting glucose  $\geq 126$   $\text{mg}/\text{dL}$ ); and hypertension (prior physician diagnosis or blood pressure  $\geq 140/90$   $\text{mmHg}$ ).

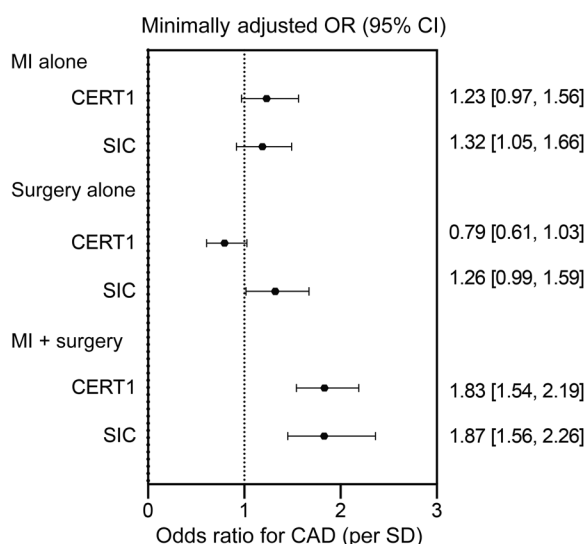
**Blood sample collection, processing, and storage.** Blood samples were collected in the morning following a 12- to 16-hour overnight fast and prepared according to the guidelines of the Lipid Research



**Figure 8. Comparison of conventional CAD risk markers with novel sphingolipid scores in the Utah CAD study.** (A) ROC curve for novel SIC score and conventional risk markers. (B) ROC curve for AHA/ACC-based clinical risk factors (age, sex, BMI, diabetes, hypertension, smoking). (C) The same AHA/ACC guidelines in addition to the CERT1 score and (D) the SIC score. ROC curves for (E) the aforementioned AHA/ACC clinical markers in addition to lipid markers (total-C, HDL-C, LDL-C, VLDL-C, TGs), (F) the clinical and lipid markers in addition to CERT1, and (G) the SIC score. For B–G, the C-statistics are indicated on the respective graphs by AUC.

Clinic's program *Manual of Laboratory Operations* (62). Lipoprotein concentrations were measured using a microscale ultracentrifugation method (63, 64). Serum samples were aliquoted and stored at  $-80^{\circ}\text{C}$ . The collection laboratory participates in the Centers for Disease Control Lipid Standardization Program (65). Of note, blood sphingolipids have been shown to be highly stable over relevant preanalytical conditions including multiple freeze-thaw cycles, temperature, long-term storage, and centrifugation time/speed (54, 55).

**Lipid extraction.** The method for conducting high-throughput lipid extraction from serum samples was modified from a method described previously (20). The internal standard (IS) stock solution containing sphingomyelin (d18:1/17:0) (2502 pmol/sample), dihydrocer (d18:0/18:1) (5 pmol/sample), d7-ceramide (d18:1-d7/16:0) (6 pmol/sample), d7-ceramide (d18:1-d7/18:0) (2 pmol/sample), d7-ceramide (d18:1/24:0) (152 pmol/sample), d7-ceramide (d18:1/24:1) (20 pmol/sample), and glucosylceramide (d18:1/17:0) (50 pmol/sample)



**Figure 9. Association of sphingolipid scores with CAD, stratified by disease presentation (MI alone, surgery alone, MI plus surgery).** OR (95% CI) for CAD per SD of sphingolipid species in the Utah CAD study, adjusted for age, sex, and BMI.

was prepared in methanol. Serum samples were thawed at 4°C for 12 hours before proceeding with lipid extraction. Samples were extracted in a 96-well format with 3 columns of controls: a 600-μL isopropanol double blank (DB), a process blank (PB) with 50 μL PBS, and a pooled control human serum sample (quality control [QC]) (MilliporeSigma). Serum (50 μL) was transferred into the remaining 72 wells of the 96-deep-well plate (USA Scientific). The IS mix (550 μL) and protein precipitation (PPT) solvent (ethyl acetate/isopropanol, 2:8, v/v) were added to each sample (with the exception of the DB) for a final volume of 600 μL per well. The plate was sealed using a silicone cap mat (Analytical Sales and Products). Samples were placed on a shaker at room temperature for 10 minutes followed by a 10-minute centrifugation at 3000 ×g. The supernatant was then transferred onto a 96-well plate (USA Scientific) and sealed with heat-sealing foil (Beckman Coulter), and plates were stored at 4°C preceding liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis.

**Lipid standards and other chemicals and reagents.** Sphingomyelin (d18:1/17:0), dihydro-cer (d18:0/18:1), d7-ceramide (d18:1-d7/16:0), d7-ceramide (d18:1-d7/18:0), d7-ceramide (d18:1-d7/24:0), d7-ceramide (d18:1-d7/24:1), and glucosylceramide (d18:1/17:0) were obtained from Avanti Polar Lipids. An Acquity CSH C18, 1.7-μm VanGuard Pre-Column and an Acquity CSH C18, 2.1 × 50 mm 1.7-μm column were obtained from Waters Corporation. 2-propanol, acetonitrile, and formic acid (all LC-MS grade) were obtained from Honeywell, Burdick & Jackson. HPLC-grade ethyl acetate was obtained from MilliporeSigma. Ammonium acetate was acquired from MPBio.

**LC-MS/MS analysis.** Lipid extracts were separated on an Acquity CSH C18 1.7 μm 2.1 × 50 mm column with a 1.7 μM VanGuard Pre-Column (Waters Corporation) maintained at 60°C and connected to an Agilent HiP 1290 Sampler and an Agilent 1290 Infinity Pump, equipped with an Agilent 1290 Flex Cube and an Agilent 6490 triple quadrupole (QqQ) mass spectrometer. Sphingolipids were detected using dynamic multiple reaction monitoring (dMRM) in positive ion mode. The source gas temperature was set to 210°C, with a gas (N<sub>2</sub>)

flow of 11 L/min and a nebulizer pressure of 30 psi. The sheath gas temperature was 400°C, the sheath gas (N<sub>2</sub>) flow was 12 L/min, the capillary voltage was 4000 V, the nozzle voltage was 500 V, high-pressure radiofrequency was 190 V, and low-pressure radiofrequency was 120 V. The injection volume was 2 μL, and the samples were analyzed in a randomized order, with the pooled QC sample injected 8 times throughout the sample queue. With 8 controls per plate, there were 80 QC injections in totality. Mobile phase A consisted of ACN/H<sub>2</sub>O (60:40 v/v), and mobile phase B consisted of IPA/ACN/H<sub>2</sub>O (90:9:1 v/v), both of which contained 10 mM ammonium formate and 0.1% formic acid. The chromatography gradient started at 15% mobile phase B, increased to 30% B over 1 minute, increased to 70% B from 1.0–1.1 minutes, was held at 70% B until 4.5 minutes, and increased to 99% B from 4.5–4.51 minutes, at which point it was held until 5 minutes, and then returned to the starting conditions at 5.1 minutes. Post-time was 1.5 minutes, and the flow rate was 0.5 mL/min throughout. Collision energies and cell accelerator voltages were optimized using sphingolipid standards with dMRM transitions as [M+H]<sup>+</sup>→[m/z = 266.3 or 284.4] for dihydroceramides; [M-H<sub>2</sub>O+H]<sup>+</sup>→[m/z = 264.2] for ceramides; and [M-H<sub>2</sub>O+H]<sup>+</sup>→[m/z = 271.3] for isotope-labeled ceramides. Sphingomyelins were monitored with dMRM transitions as [M+H]<sup>+</sup>→[m/z = 184.4]. Sphingolipids without available standards were identified on the basis of high-resolution LC-MS, quasi-molecular ions, and characteristic product ions. Results from the LC-MS experiments were collected using an Agilent Mass Hunter Workstation and analyzed with Agilent Mass Hunter Quant B.07.00 software. Sphingolipids were quantitated on the basis of peak area ratios to the internal standards.

**Lipid species.** A total of 32 lipids were quantified including dihydroceramides [dihydro-cer(d18:0)]; ceramides [cer(d18:1)]; glucosylceramides [glucosyl-cer(d18:1)]; dihydrosphingomyelins [dihydro-SM(d18:0)]; sphingomyelins [SM(d18:1)]; sphinganine; and sphingosine. For each of these, except for sphinganine and sphingosine, acyl chain lengths of 16, 18, 20, 22, and 24 and a carbon length of 24:1 were reported. The median (IQR) coefficient of variation (11.76, 6.85–20.53) was comparable to previously published sphingolipid data (66).

To calculate the CERT1 score that is in clinical use (50), we calculated C16:0, C18:0, and C24:1 concentrations and their ratio to C24:0, assigning 2 points to those with levels in the fourth quartile, 1 point to the third quartile, and 0 points to the bottom 2 quartiles, with total CERT1 scores ranging from 0–12 (22).

**Table 5. Stratification of Utah CAD participants by disease severity**

	Controls	Surgery alone	MI alone	MI + surgery
n (%)	212 (100%)	75 (16%)	82 (18%)	305 (66%)
CERT1	3.24 ± 2.7	3.9 ± 2.5	5 ± 3.3	5.4 ± 3.2
SIC	5.5 ± 2.4	10.7 ± 9.3	10.8 ± 7.5	12.2 ± 12.5
LDL-C	103.1 ± 2.8	112 ± 37.4	142.9 ± 47.6	128.9 ± 40.2
Total-C	189.9 ± 3.3	192 ± 43.4	220.6 ± 49.1	210.2 ± 44.4
TGs	178 ± 9.5	193 ± 133.2	175.8 ± 89.3	212.2 ± 152.3

Clinical lipid marker serum concentrations were stratified by disease severity and are presented as concentration (mg/dL) ± SD. Surgery denotes PTCA or CABG.



**Statistics.** Participant characteristics were summarized as the mean  $\pm$  SD for continuous variables or N (percentage) for categorical variables (Table 1). Differences between patients and controls were compared using a 2-tailed Student's *t* test for continuous variables and a  $\chi^2$  test for categorical variables. *P* values of less than 0.05 were considered significant. Lipid species were summarized as medians and IQRs using the original scale (Table 2) and were  $\log_{10}$  transformed for analysis, owing to non-normal distribution. When assessing the effect of summed molecular lipid species or acyl chains on CAD, the variables were summed preceding log transformation.

Multivariable-adjusted and unadjusted ORs and 95% CIs were estimated using logistic regression and reported per the SD (of lipid species). A priori-defined covariates based on current ACC and AHA guidelines were considered in stepwise variable selection modeling (Supplemental Table 3). These covariates included the following: age, sex, BMI, total-C, LDL-C, HDL-C, VLDL-C, TGs, hypertension, diabetes, and smoking. We calculated the percentage of change in the OR from the parsimonious age-, sex-, and BMI-adjusted model with the addition of each covariate, though no covariate affected all sphingolipids. Our final parsimonious model included age, sex, and BMI, but we also show results for a fully adjusted model including all AHA/ACC guideline-based risk factors in the main figures for comparison. In addition to testing whether AHA/ACC-based risk factors were confounders of the sphingolipid-CAD relationship, we evaluated some of these variables for potential effect modification through the inclusion of a variable by lipid interaction term in the logistic regression models and evaluation of the significance of the interaction term using a likelihood ratio test (Supplemental Table 4 and ref. 67). Where effect modification was present (*P* for the interaction term  $< 0.05$ ; Supplemental Table 4), we ran the analyses separately according to levels of the effect modifier (e.g., hypertensive or normotensive) to determine whether the relationships between sphingolipids and CAD differed according to subgroups of the effect modifier variable (Supplemental Table 4).

We applied machine learning (24, 25) to identify the most predictive biomarkers. To compare classical variable reduction techniques with our machine-learning approaches, we performed a stepwise (forward and backward) regression (Supplemental Table 5). We then performed LASSO regression (Supplemental Table 6 and ref. 68) and RF analysis (Supplemental Table 7 and ref. 69). AHA/ACC lipid risk factor variables (LDL-C, etc.) were included along with the sphingolipids as input variables to allow the machine-learning algorithm to determine the most predictive lipid biomarkers. Data were split into training (80%) and testing (20%) data sets. For LASSO analysis, the optimal value for the tuning parameter  $\lambda$  was selected to maximize the percentage of correctly identified cases/controls with 10-fold cross validation on the training set before using the remaining 20% of the data to test the predictability of the model. We determined the quality of prediction via the percentage of correctly identified cases/controls, averaging the percentages across 10 training and testing splits. There were 2 data input approaches for RF analysis. For a sphingolipid-only

input, 32 sphingolipid variables were used, with a default of 500 decision trees to generate an optimal number of variables per tree determined for each of 5 cross-validation training sets. Variable importance scores were assigned through permutation testing, and the top 5 variables averaged across validation sets were placed into a single model. A second input included the 32 sphingolipid variables and classical CAD markers (i.e., cholesterol, TGs, etc.). To examine conditional correlations ( $r \geq 0.20$ ) between ceramides and conventional biomarkers in CAD cases, we generated a GGM with visualization in Cytoscape (70–75). GGMs model conditional dependencies among continuous variables with multivariate Gaussian distributions. Recent studies have demonstrated how GGMs, which are data driven, can reconstruct biological pathway reactions (76). We generated a GGM to determine whether our sphingolipid panel was redundant in the presence of traditional clinical lipid biomarkers (i.e., whether they were highly correlated, conditioned on the presence of all other lipids).

For comparison of the ability of different clinical markers and scores to distinguish between true cases and controls, we applied ROC-AUC analysis and calculated the NRI and IDI (Supplemental Table 1 and ref. 77). The ROC curves and C-statistics are presented in Supplemental Figure 1.

All analyses were performed in R 3.5.1 (78). Associations were considered statistically significant at a FDR below 0.05 to control for multiple statistical tests (Supplemental Table 8).

**Study approval.** Lipid quantification and secondary data analysis of these patient samples were approved by the IRB of the University of Utah, and all patients provided written informed consent.

## Author contributions

AMP, BJH, MCP, WLH, and SAS conceived the project, designed the experiments, processed the ceramide data, and wrote the manuscript. AMP, JAM, and JEC conducted the lipidomic analyses. PNH and SCH developed the CAD cohort including recruitment of patients, collection of clinical and demographic data, and storing of samples.

## Acknowledgments

We are grateful for the support from the Metabolomics Core at the Health Sciences Center of the University of Utah. The authors received research support from the NIH (DK112826 and DK108833, to WLH, and DK115824, DK116888, and DK116450, to SAS); the JDRF (3-SRA-2019-768-A-B, to WLH); the American Diabetes Association (to SAS); the AHA (to SAS); the Margolis Foundation (to SAS); the National Cancer Institute (5R00CA218694-03, to MCP); and the Huntsman Cancer Institute Cancer Center (P30CA040214, to MCP).

Address correspondence to: Scott Summers, 15N 2030E, Salt Lake City, Utah 84112, USA. Phone: 801.585.9552; Email: scott.a.summers@health.utah.edu.

1. Rautou P-E, et al. Microparticles, vascular function, and atherothrombosis. *Circ Res*. 2011;109(5):593–606.
2. Bergman BC, et al. Serum sphingolipids: relationships to insulin sensitivity and changes with exercise in humans. *Am J Physiol Endocrinol Metab*.

2015;309(4):E398–E408.

3. AHA. Cardiovascular Disease Burden Report. Cardiovascular disease: a costly burden for America. <http://www.heart.org/en/get-involved/advocate/federal-priorities/cardiovascular-disease-burden-report>. Accessed December 4, 2019.

4. Summers SA. Could ceramides become the new cholesterol? *Cell Metab*. 2018;27(2):276–280.
5. Summers SA, Chaurasia B, Holland WL. Metabolic messengers: ceramides. *Nat Metab*. 2019;1:1051–1058.
6. Merrill AH. De novo sphingolipid biosynthesis: a



- necessary, but dangerous, pathway. *J Biol Chem.* 2002;277(29):25843–25846.
7. Chaurasia B, et al. Targeting a ceramide double bond improves insulin resistance and hepatic steatosis. *Science.* 2019;365(6451):386–392.
  8. Zelnik ID, Rozman B, Rosenfeld-Gur E, Ben-Dor S, Futerman AH. A stroll down the CerS lane. *Adv Exp Med Biol.* 2019;1159:49–63.
  9. Hammerschmidt P, et al. CerS6-derived sphingolipids interact with Mff and promote mitochondrial fragmentation in obesity. *Cell.* 2019;177(6):1536–1552.e23.
  10. Raichur S, et al. The role of C16:0 ceramide in the development of obesity and type 2 diabetes: CerS6 inhibition as a novel therapeutic approach. *Mol Metab.* 2019;21:36–50.
  11. Turpin-Nolan SM, et al. CerS1-derived C<sub>18:0</sub> ceramide in skeletal muscle promotes obesity-induced insulin resistance. *Cell Rep.* 2019;26(1):1–10.e7.
  12. Turpin SM, et al. Obesity-induced CerS6-dependent C16:0 ceramide production promotes weight gain and glucose intolerance. *Cell Metab.* 2014;20(4):678–686.
  13. Hla T, Kolesnick R. C16:0-ceramide signals insulin resistance. *Cell Metab.* 2014;20(5):703–705.
  14. Raichur S, et al. CerS2 haploinsufficiency inhibits  $\beta$ -oxidation and confers susceptibility to diet-induced steatohepatitis and insulin resistance. *Cell Metab.* 2014;20(4):687–695.
  15. Chaurasia B, Summers SA. Ceramides — lipotoxic inducers of metabolic disorders. *Trends Endocrinol Metab.* 2015;26(10):538–550.
  16. de Carvalho LP, et al. Plasma ceramides as prognostic biomarkers and their arterial and myocardial tissue correlates in acute myocardial infarction. *JACC Basic Transl Sci.* 2018;3(2):163–175.
  17. Hilvo M, et al. Development validation of a ceramide-phospholipid-based cardiovascular risk estimation score for coronary artery disease patients [published online ahead of print June 18, 2019]. *Eur Heart J.* <https://doi.org/10.1093/eurheartj/ehz387>.
  18. Laaksonen R, et al. Plasma ceramides predict cardiovascular death in patients with stable coronary artery disease and acute coronary syndromes beyond LDL-cholesterol. *Eur Heart J.* 2016;37(25):1967–1976.
  19. Peterson LR, et al. Ceramide remodeling and risk of cardiovascular events and mortality. *J Am Heart Assoc.* 2018;7(10):e007931.
  20. Havulinna AS, et al. Circulating ceramides predict cardiovascular outcomes in the population-based FINRISK 2002 cohort. *Arterioscler Thromb Vasc Biol.* 2016;36(12):2424–2430.
  21. Kaasenbrood L, et al. Distribution of estimated 10-year risk of recurrent vascular events and residual risk in a secondary prevention population. *Circulation.* 2016;134(19):1419–1429.
  22. Kauhanen D, Sysi-Aho M, Koistinen KM, Laaksonen R, Sinisalo J, Ekroos K. Development and validation of a high-throughput LC-MS/MS assay for routine measurement of molecular ceramides. *Anal Bioanal Chem.* 2016;408(13):3475–3483.
  23. Gotto AM. Interrelationship of triglycerides with lipoproteins and high-density lipoproteins. *Am J Cardiol.* 1990;66(6):20A–23A.
  24. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods.* 2018;15(4):233–234.
  25. Song X, Mitnitski A, Cox J, Rockwood K. Comparison of machine learning techniques with classical statistical models in predicting health outcomes. *Stud Health Technol Inform.* 2004;107(Pt 1):736–740.
  26. Massey KJ, Hong NJ, Garvin JL. Angiotensin II stimulates superoxide production in the thick ascending limb by activating NOX4. *Am J Physiol, Cell Physiol.* 2012;303(7):C781–C789.
  27. Pencina MJ, D'Agostino RB, Pencina KM, Janssens AC, Greenland P. Interpreting incremental value of markers added to risk prediction models. *Am J Epidemiol.* 2012;176(6):473–481.
  28. Hall SL, Lorenc T. Secondary prevention of coronary artery disease. *Am Fam Physician.* 2010;81(3):289–296.
  29. Anroedh S, et al. Plasma concentrations of molecular lipid species predict long-term clinical outcome in coronary artery disease patients. *J Lipid Res.* 2018;59(9):1729–1737.
  30. Mantovani A, et al. Association of plasma ceramides with myocardial perfusion in patients with coronary artery disease undergoing stress myocardial perfusion scintigraphy. *Arterioscler Thromb Vasc Biol.* 2018;38(12):2854–2861.
  31. Hilvo M, et al. PCSK9 inhibition alters the lipidome of plasma and lipoprotein fractions. *Atherosclerosis.* 2018;269:159–165.
  32. Meeusen JW, Donato LJ, Bryant SC, Baudhuin LM, Berger PB, Jaffe AS. Plasma ceramides. *Arterioscler Thromb Vasc Biol.* 2018;38(8):1933–1939.
  33. Lemaitre RN, et al. Circulating sphingolipids, insulin, HOMA-IR, and HOMA-B: The Strong Heart Family Study. *Diabetes.* 2018;67(8):1663–1672.
  34. Wigger L, et al. Plasma dihydroceramides are diabetes susceptibility biomarker candidates in mice and humans. *Cell Rep.* 2017;18(9):2269–2279.
  35. Jensen PN, et al. Circulating sphingolipids, fasting glucose, and impaired fasting glucose: The Strong Heart Family Study. *EBioMedicine.* 2019;41:44–49.
  36. Lee SY, et al. Cardiomyocyte specific deficiency of serine palmitoyltransferase subunit 2 reduces ceramide but leads to cardiac dysfunction. *J Biol Chem.* 2012;287(22):18429–18439.
  37. Park TS, Goldberg JJ. Sphingolipids, lipotoxic cardiomyopathy, and cardiac failure. *Heart Fail Clin.* 2012;8(4):633–641.
  38. Park TS, et al. Ceramide is a cardiotoxin in lipotoxic cardiomyopathy. *J Lipid Res.* 2008;49(10):2101–2112.
  39. Ussher JR, et al. Inhibition of de novo ceramide synthesis reverses diet-induced insulin resistance and enhances whole-body oxygen consumption. *Diabetes.* 2010;59(10):2453–2464.
  40. Glaros EN, Kim WS, Quinn CM, Jessup W, Rye KA, Garner B. Myriocin slows the progression of established atherosclerotic lesions in apo-lipoprotein E gene knockout mice. *J Lipid Res.* 2008;49(2):324–331.
  41. Glaros EN, et al. Inhibition of atherosclerosis by the serine palmitoyl transferase inhibitor myriocin is associated with reduced plasma glycosphingolipid concentration. *Biochem Pharmacol.* 2007;73(9):1340–1346.
  42. Hojati MR, et al. Effect of myriocin on plasma sphingolipid metabolism and atherosclerosis in apoE-deficient mice. *J Biol Chem.* 2005;280(11):10284–10289.
  43. Park TS, Rosebury W, Kindt EK, Kowala MC, Panek RL. Serine palmitoyltransferase inhibitor myriocin induces the regression of atherosclerotic plaques in hyperlipidemic ApoE-deficient mice. *Pharmacol Res.* 2008;58(1):45–51.
  44. Holland WL, et al. Inhibition of ceramide synthesis ameliorates glucocorticoid-, saturated-fat-, and obesity-induced insulin resistance. *Cell Metab.* 2007;5(3):167–179.
  45. Park TS, et al. Modulation of lipoprotein metabolism by inhibition of sphingomyelin synthesis in ApoE knockout mice. *Atherosclerosis.* 2006;189(2):264–272.
  46. Bharath LP, et al. Ceramide-initiated protein phosphatase 2A activation contributes to arterial dysfunction in vivo. *Diabetes.* 2015;64(11):3914–3926.
  47. Zhang QJ, et al. Ceramide mediates vascular dysfunction in diet-induced obesity by PP2A-mediated dephosphorylation of the eNOS-Akt complex. *Diabetes.* 2012;61(7):1848–1859.
  48. Bikman BT, et al. Fenretinide prevents lipid-induced insulin resistance by blocking ceramide biosynthesis. *J Biol Chem.* 2012;287(21):17426–17437.
  49. Holland WL, et al. Lipid-induced insulin resistance mediated by the proinflammatory receptor TLR4 requires saturated fatty acid-induced ceramide biosynthesis in mice. *J Clin Invest.* 2011;121(5):1858–1870.
  50. Vasilakaité-Brooks I, et al. Structural insights into adiponectin receptors suggest ceramidase activity. *Nature.* 2017;544(7648):120–123.
  51. Holland WL, et al. Receptor-mediated activation of ceramidase activity initiates the pleiotropic actions of adiponectin. *Nat Med.* 2011;17(1):55–63.
  52. Holland WL, et al. An FGF21-adiponectin-ceramide axis controls energy expenditure and insulin action in mice. *Cell Metab.* 2013;17(5):790–797.
  53. Weng SF, Reys J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS ONE.* 2017;12(4):e0174944.
  54. Wagner-Golbs A, et al. Effects of long-term storage at  $-80^{\circ}\text{C}$  on the human plasma metabolome. *Metabolites.* 2019;9(5):E99.
  55. Brunkhorst R, et al. Preanalytical biases in the measurement of human blood Sphingolipids. *Int J Mol Sci.* 2018;19(5):E1390.
  56. Aronson JK, Ferner RE. Biomarkers — a general review. *Curr Protoc Pharmacol.* 2017;76:9.23.1–9.23.17.
  57. Hopkins PN, Wu LL, Hunt SC, Brinton EA. Plasma triglycerides and type III hyperlipidemia are independently associated with premature familial coronary artery disease. *J Am Coll Cardiol.* 2005;45(7):1003–1012.
  58. Hunt SC, Williams RR, Barlow GK. A comparison of positive family history definitions for defining risk of future disease. *J Chronic Dis.* 1986;39(10):809–821.
  59. Vikulova DN, et al. Premature atherosclerotic cardiovascular disease: trends in incidence, risk factors, and sex-related differences, 2000 to 2016. *J Am Heart Assoc.* 2019;8(14):e012178.
  60. Higgins M, et al. NHLBI Family Heart Study: objectives and design. *Am J Epidemiol.* 1996;143(12):1219–1228.

61. Hopkins PN, et al. Altered composition of ceramide-rich lipoproteins and coronary artery disease in a large case-control study. *Atherosclerosis*. 2009;207(2):559–566.
62. National Heart and Lung Institute (US), Lipid Research Clinics Program. *Manual of Laboratory Operations, Lipid and Lipoprotein Analysis*, vol. 1. National Heart and Lung Institute, National Institutes of Health, Washington, DC, USA; 1974.
63. Warnick GR, Benderson J, Albers JJ. Dextran sulfate-Mg<sup>2+</sup> precipitation procedure for quantitation of high-density-lipoprotein cholesterol. *Clin Chem*. 1982;28(6):1379–1388.
64. Wu LL, Warnick GR, Wu JT, Williams RR, Lalouel JM. A rapid micro-scale procedure for determination of the total lipid profile. *Clin Chem*. 1989;35(7):1486–1491.
65. Stratford S, Hoehn KL, Liu F, Summers SA. Regulation of insulin action by ceramide: dual mechanisms linking ceramide accumulation to the inhibition of Akt/protein kinase B. *J Biol Chem*. 2004;279(35):36608–36615.
66. Sullards MC, Liu Y, Chen Y, Merrill AH. Analysis of mammalian sphingolipids by liquid chromatography tandem mass spectrometry (LC-MS/MS) and tissue imaging mass spectrometry (TIMS). *Biochim Biophys Acta*. 2011;1811(11):838–853.
67. Zeileis A, Hothorn T. Diagnostic checking in regression relationships. *R News*. 2002;2(3):7–10.
68. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1–22.
69. Liaw A, Weiner M. Classification and regression by random forest. *R News*. 2002;2(3):18–22.
70. Shannon P, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–2504.
71. Butts CT. Network: classes for relational data. <https://cran.r-project.org/package=network>. Accessed December 4, 2019.
72. Butts CT. Network: a package for managing relational data in R. *J Stat Softw*. 2008;24(2):1–36.
73. Genz A, Bretz F. *Computation of Multivariate Normal and T Probabilities*. Berlin, Heidelberg, Germany: Springer Verlag; 2009.
74. Genz A, et al. mvtnorm: multivariable normal and t distributions. <http://mvtnorm.r-forge.r-project.org>. Updated: June 19, 2019. Accessed December 4, 2019.
75. Friedman J, Hastie T, Tibshirani J. Glasso: graphical LASSO: estimation of Gaussian graphical models. <https://cran.r-project.org/package=glasso>. Accessed December 4, 2019.
76. Krumsiek J, Suhre K, Illig T, Adamski J, Theis FJ. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC Syst Biol*. 2011;5:21.
77. Robin X, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
78. R Software. R Development Core Team. R Foundation for Statistical Computing. 2018.