JCI The Journal of Clinical Investigation

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. 2019. https://doi.org/10.1172/JCI131838.

Clinical Research and Public Health In-Press Preview **Metabolism** **Vascular biology**

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 of individuals with familial coronary artery disease (CAD) (n = 462) and population-based controls (n = 212) to explore the relationship between serum sphingolipids and CAD, employing 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 population controls. We generated a novel Sphingolipid Inclusive CAD risk score, termed SIC, that demarcates CAD patients independently and more effectively than conventional clinical CVD biomarkers including LDL-cholesterol and serum triglycerides. This new metric comprises several minor lipids which likely serve as measures of flux through the ceramide biosynthesis pathway, rather than the abundant deleterious ceramide species that are incorporated in other ceramide-based scores.

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

Machine Learning Reveals Serum Sphingolipids as Cholesterol-Independent Biomarkers of Coronary Artery Disease

Authors: Annelise M. Poss[1], J. Alan Maschek[2], James E. Cox[2], Benedikt J. Hauner[3,4], Paul N. Hopkins[5], Steven C. Hunt[5,6], William L. Holland[1], Scott A. Summers[1*†] and Mary C. Playdon[1,3†]

Affiliations:
1. Department of Nutrition and Integrative Physiology and the Diabetes and Metabolism Research Center, University of Utah, Salt Lake City, UT 84112
2. Department of Biochemistry and the Metabolomics and Proteomics Core Research Facilities, University of Utah, Salt Lake City, UT 84112
3. Division of Cancer Population Sciences, Huntsman Cancer Institute, Salt Lake City, UT 84112
4. Department of Mathematics, Technical University of Munich, Munich, Germany 85748
5. Department of Internal Medicine, University of Utah, Salt Lake City, UT 84112
6. Department of Genetic Medicine, Weill Cornell Medicine, Doha, Qatar

*Correspondence to: Scott Summers, 15N 2030E, Salt Lake City, UT 84112,
Ph: 801-585-9552; Email: scott.a.summers@health.utah.edu

†These authors contributed equally to this study

Conflict of Interests: S.A.S is a cofounder, consultant, and shareholder for Centaurus Therapeutics.

ABSTRACT

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 of individuals with familial coronary artery disease (CAD) (n=462) and population-based controls (n=212) to explore the relationship between serum sphingolipids and CAD, employing 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 population controls. We generated a novel Sphingolipid Inclusive CAD risk score, termed SIC, that demarcates CAD patients independently and more effectively than conventional clinical CVD biomarkers including LDL-cholesterol and serum triglycerides. This new metric comprises several minor lipids which likely serve as measures of flux through the ceramide biosynthesis pathway, rather than the abundant deleterious ceramide species that are incorporated in other ceramide-based scores.

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

Funding: National Institutes of Health (DK112826 and DK108833 to WLH and DK115824, DK116888, and DK116450 to SAS), the Juvenile Diabetes Research Foundation (JDRF 3-SRA-2019-768-A-B to WLH), the American Diabetes Association (to SAS), the American Heart Association (to SAS), the Margolis Foundation (to SAS), the National Cancer Institute (5R00CA218694-03 to MCP), and the Huntsman Cancer Institute Cancer Center Support Grant (P30CA040214 to MCP).

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). Family history of CAD is an independent risk factor for MI, and once a patient has undergone an MI they are at greatly increased risk for subsequent adverse cardiac events. In addition to incurring a substantial individual health burden, CVD is the United States' costliest disease, producing an economic toll that is projected to grow substantially over the coming decades (3). The combination of personal and financial costs necessitates development of 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 cardiovascular health include triglycerides and cholesterol, two abundant and easily quantifiable circulating factors. Recent technological advances now allow for detection of less plentiful lipids, such as sphingolipids, enabling substantially more diverse lipidomic screenings at relatively high throughput. 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 cardiovascular disease indicates that these sphingolipids play causative roles in diabetes and cardiometabolic disorders (5).

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 (5)(*Figure 1*), with the primary steps of *de novo* synthesis occurring in the endoplasmic reticulum (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 retention of lipoproteins in the vascular wall, impaired vasodilation, and induction of cardiomyocyte apoptosis (15).

Prior profiling studies have identified three ceramide species [i.e. cer(d18:1/16:0), cer(d18:1/18:0), and cer(d18:1/24:0)] that positively associate with CVD incidence (16), secondary CVD events (17) and mortality (18-20). Cer(d18:1/24:0) has been reported to negatively associate 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 biosynthetic pathway. Therefore, we performed an inclusive sphingolipid screen (32 sphingolipids) in individuals with CAD and 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 coronary artery disease.

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, excepting two glucosylceramides, were elevated in CAD cases compared with controls (*Table 2*). Sphingosine (p-value < 2E-16), dihydro-cer(d18:0/ 16:0) (p-

value < 2E-16), dihydro-cer(d18:0/ 18:0) (p-value < 2E-16), and cer(d18:1/ 24:1) (p-value < 2E-16) were most strongly associated with CAD (OR_{perSD} 3.47, 95% CI 2.63-4.69; OR_{perSD} 2.54, 95% CI 2.06-3.18; OR_{perSD} 2.82, 95% CI 2.24-3.60; OR_{perSD} 2.30, 95% CI 2.24-3.60; OR_{perSD} 2.29, 95% CI 1.86, 2.85, respectively). *Figure 3* depicts the odds ratios (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, BMI), and a fully-adjusted model (i.e. a model that includes the covariates age, sex, body mass index (BMI), total cholesterol, LDL cholesterol, HDL cholesterol, VLDL cholesterol, triglycerides, 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 cardiovascular mortality (4, 21, 22). CERT1 performed well in this cohort, as subjects with CAD had a significantly higher CERT1 risk scores compared with control participants (OR_{perSD} 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_{perSD} 2.30, 95% CI 1.87-2.=]6; cer(d18:1/18:0); OR_{perSD} 2.30, 95% CI 1.87-2.85; cer(d18:1/24:1); OR_{perSD} 2.29, 95% CI 1.86-2.85] (*Figure3*). Since cer(d18:1/24:0) was also elevated in individuals with CAD (OR_{perSD} 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 two different categories. In one 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_{perSD} 2.66, CI 2.12-3.38) and/or total dihydroceramides, independent of chain length (OR_{perSD} 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, triglycerides, 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 inter-connected 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 triglycerides (23). Ceramides did not correlate with VLDL-C, triglycerides, 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 employed 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

recognition within complex datasets 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 non-linear (24, 25). Moreover, machine learning generated models tend to me 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*). RF develops algorithms that can precisely classify observations into groups (i.e. CAD cases versus controls). With this method, the number of variables incorporated has a strong impact on model accuracy: if variables improve model fit, RF accuracy improves; if not, accuracy is diluted by meaningless variables. We therefore ran two 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, triglycerides). For our LASSO approach, the input included all sphingolipids and the aforementioned conventional CVD lipid markers. Biomarker score classification was evaluated using both ORs and receiver operator characteristic-area under the curve (ROC-AUC) analysis (*Figure 7, Table 4*). For both RF and LASSO approaches, the five lipids most positively associated with CAD were used to generate a score.

An RF-generated Sphingolipid Inclusive CAD Risk Score (RF-SIC) (AUC = 0.75) outperformed CERT1 (AUC = 0.67) and conventional CVD risk biomarkers including LDL-C (AUC = 0.69) and total cholesterol (AUC = 0.63) (*Figure 7*, Supplemental Figure 1). An RF model generated from the sphingolipids plus CVD risk markers (denoted with an asterisk, RF-SIC^{+}, AUC = 0.78) included LDL-C and displayed more precise classification of CAD cases versus controls, as compared to the RF-SIC score that excluded LDL-C (*Figure 7*). When

evaluated by OR, RF-SIC^{+} (OR 5.03, 95% CI 3.69-7.07) outperformed RF-SIC (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 vs. 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*, *Supplemental Figure 2*). Adding in another ratio (i.e. the second highest, positively-associated lipid vs. 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, Supplemental Figure 2*).

Based on 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, Table 4*). Only sphingolipids, and not LDL-C, were included in the final SIC score so comparisons of an inclusive sphingolipid measure to conventional CVD lipid markers could be performed. A list of the lipid components in each novel score can be found in *Table 3*.

Comparing Machine Learning-Generated Scores to Conventional Markers of CAD

We next compared the ability of SIC, CERT1, and standard clinical biomarkers (triglycerides, LDL-C, etc.) to classify CAD cases compared with controls (*Figures 8A, Table 4)*. We provide the following ROC curves (with the area under the curve, AUC) for comparison (*Figures 8B to 8G*): 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 AUC can be

an insensitive measure of model performance, particularly when the initial model (i.e. the American Heart Association (AHA)/American College of Cardiology (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-value <0.0001) and IDI was 0.10 (95% CI: 0.08-0.11, p-value <0.0001) (*Supplemental Table 1*). For frame of reference, an NRI exceeding 0.6 is considered strong and 0.4 is considered intermediate (27). SIC was superior to CERT1, which had an NRI of 0.48 (95% CI: 0.32-0.64, p-value <0.0001) and IDI of 0.04 (95% CI: 0.03-0.06, p-value <0.0001) (*Supplemental Table 1*). 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), 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 while 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 CAD patients that were stratified into three subgroups: (a) patients having had a myocardial infarction (MI) alone; (b) patients that had a surgical intervention alone (coronary artery bypass grafting (CABG) or percutaneous transluminal angioplasty (PCTA)); or, (c) patients that had an MI in combination with a surgical intervention. Patients undergoing a surgical intervention alone are considered to have a more tightly controlled disease state, while those with both surgical interventions and

MIs are likely to be in a more severe or uncontrolled disease state (28). Patients with an MI alone are considered intermediate. As compared to 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_{perSD} > 1.80, p<5 $x10^{-11}$; p-heterogeneity<2E-16, *Figure 9a*). By comparison, standard clinical markers including LDL-C, total-C, and triglyceride didn't 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 subjects with CAD compared with healthy controls. Thirty of the thirty-two 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 which we have termed SIC (i.e. sphingolipid inclusive CAD risk score) that 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 from the regression output, then log transformed. This score approached a strong C-statistic of 0.79 and an OR_{perSD} of 4.67 (95% CI: 3.46-6.43) for risk of CAD, outperforming other serum indices of cardiovascular risk including LDL-C alone and the CERT1 ceramide risk score. Serum ceramides 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). They were also increased in individuals with insulin resistance or type 2 diabetes (33-35), two underlying drivers of cardiovascular morbidity (33, 34). These studies implicate ceramide as a marker of disease pathology, disease risk, mortality, and 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, CERT1's 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 dataset as compared to previous prospective cohort studies, 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 dataset. However, SIC and CERT1 showed stronger discriminatory power than LDL-C as 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 triglycerides. . 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 cardiovascular 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 likely causative agents in disease progression (15). Studies in rodent models reveal 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 which underlie CVD. Manipulations of the de novo ceramide synthesis pathway further suggest that certain ceramide species are deleterious while 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. Lastly, studies in rodents reveal that ceramide degradation is a primary means by which adiponectin receptors, which are ligand-activated ceramidases (50), exert their anti-diabetic, 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, dihydro-cer(d18:0/16:0), dihydro-cer(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 associated 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, 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 early-onset CAD patients (average age of onset = 47.8), 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 accuracy of models and reduce dimensionality of datasets (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). While these are not ideal, they are comparable to 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 datasets. 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, cases and controls were collected and handled in the same manner, so relative differences (and calculated odds ratios for CAD) should be sustained. Moreover, prior studies have shown that sphingolipids remain stable over storage periods as long as 16 years-post 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, they support the development of therapeutics targeting sphingolipids as a means of ameliorating cardiovascular risk. 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 in predicting risk of cardiovascular disease.

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 in Utah, USA (n=462 cases and n=212 controls)(57).

Study Population

Cases were recruited between 1990 and 2000 from Intermountain Healthcare discharge records or the Family Health Tree Program in Utah (58). Cases were aged 30-75 years with a diagnosis of CAD, defined by the original study recruitment criteria as myocardial infarction (MI), percutaneous transluminal coronary angioplasty (PCTA), or coronary artery bypass grafting (CABG). A large proportion of cases were male (77%), likely because premature CAD incidence rates are higher for men than women (59). Cases had similar age of onset to at least one 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 six months following their event.

Controls representative of the Utah population (57, 60, 61) were randomly sampled from 1980-1986 from (i) the parents of students participating in the Family Health Tree Program (58),

a study of family health among Utah high schools; and (ii) spouse pairs participating in a study on psychological factors concerning CAD (61). Control participants were aged 30-75 years and had no clinical diagnosis of CAD, but they could have a family history of CAD. Controls taking vasoconstrictive drugs (i.e. beta blockers, calcium channel blockers, and other anti-anginal medications) were excluded.

Both case and control populations were selected from the same source population of Salt Lake City, Utah. The number of cases (462) is larger than controls (212) due to 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 analyses included age (years), sex (male or female), body mass index (BMI, Kg/m^2), smoking ("ever" or "never" to smoking daily for a year or more), total cholesterol (mg/dL), LDL-cholesterol (mg/dL), VLDL-cholesterol (mg/dL), HDL-cholesterol (mg/dL), triglycerides (mg/dL), lipid medication (statins, fibrates, and other hyperlipidemia managing drugs taken at time of blood draw, yes/no), diabetes (prior physician diagnosis or fasting glucose ≥126 mg/dL), and hypertension (prior physician diagnosis or blood pressure ≥140/90 mm Hg).

Blood Sample Collection, Processing and Storage

Blood samples were collected in the morning following a 12-16 hour overnight fast and prepared according to guidelines of the Lipid Research 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 °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), dihydroceramide (d18:0/18:1) (5 pmol/sample), d_7-ceramide (d18:1-d7/16:0) (6 pmol/sample), d_7-ceramide (d18:1-d7/18:0) (2 pmol/sample), d_7-ceramide (d18:1/24:0) (152 pmol/sample), d_7-ceramide (d18:1/24:1) (20 pmol/sample), and glucosylceramide (d18:1/17:0) (50 pmol/sample) 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 three columns of controls: a 600 μl isopropanol double blank (DB), a process blank (PB) with 50 μl phosphate buffered saline (PBS), and a pooled control human serum sample (QC) (Millipore Sigma). 50 μl of serum was transferred to the remaining 72 wells of the 96-deepwell plate (USA Scientific). 550 μl IS mix and protein precipitation (PPT) solven (ethyl acetate: isopropanol (2:8, *v/v*)) was 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 3000x*g*. The supernatant was then transferred to a 96-well plate (USA Scientific) and sealed with heat-sealing foil (Beckman Coulter) and plates were stored at 4 °C preceding LC-MS/MS analysis.

Lipid Standards and Other Chemicals and Reagents

Sphingomyelin (d18:1/17:0), dihydroceramide (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, Inc. (Alabaster, AL). Acquity CSH C18, 1.7 µm VanGuard Pre-Column, and Acquity CSH C18, 2.1 x 50 mm 1.7 µm were obtained from Waters (Milford, MA). 2-propanol, acetonitrile, and formic acid (all LC-MS grade) were attained from Honeywell – Burdick & Jackson (Muskegon, MI). HPLC grade ethyl acetate was from EMD Millipore (Billerica, MA). Ammonium acetate was acquired from MPBio (Santa Ana, CA).

LC-MS/MS analysis

Lipid extracts were separated on an Acquity CSH C18 1.7 µm 2.1 x 50 mm column with a 1.7 µM VanGuard Pre-Column (Waters Corp, Milford, MA) maintained at 60 °C connected to an Agilent HiP 1290 Sampler, Agilent 1290 Infinity pump, equipped with an Agilent 1290 Flex Cube and Agilent 6490 triple quadrupole (QqQ) mass spectrometer. Sphingolipids were detected using dynamic multiple reaction monitoring (dMRM) in positive ion mode. Source gas temperature was set to 210°C, with a gas (N_2) flow of 11 L/min and a nebulizer pressure of 30 psi. Sheath gas temperature was 400°C, sheath gas (N_2) flow was 12 L/min, capillary voltage was 4000 V, nozzle voltage 500 V, high pressure RF 190 V and low-pressure RF was 120 V. Injection volume was 2 µL and the samples were analyzed in a randomized order with the pooled QC sample injected eight times throughout the sample queue. With 8 controls per plate, there were 80 QC injections in totality. Mobile phase A consisted of ACN: H_2O (60:40 *v/v*) and mobile phase B consisted of IPA: ACN: H_2O (90:9:1 *v/v*) both containing 10 mM ammonium formate and 0.1% formic acid. The chromatography gradient started at 15% mobile phase B, increased to 30% B over 1 min, increased to 70% B from 1.0-1.1 min, held at 70% B to 4.5 min, and increased to 99% B from 4.5-4.51 min where it was held until 5 min then returned to starting conditions at 5.1 min. Post-time was 1.5 min and the flowrate was 0.5 mL/min throughout.

Collision energies and cell accelerator voltages were optimized using sphingolipid standards with dMRM transitions as $[M+H]^+ \rightarrow [m/z = 266.3 \text{ or } 284.4]$ for dihydroceramides, $[M-H_2O+H]^+ \rightarrow [m/z = 264.2]$ for ceramides, $[M-H_2O+H]^+ \rightarrow [m/z = 271.3]$ for isotope labeled ceramides. Sphingomyelins were monitored with dMRM transitions as $[M+H]^+ \rightarrow [m/z = 184.4]$. Sphingolipids without available standards were identified based on HR-LC/MS, quasi-molecular ion and characteristic product ions. Results from LC-MS experiments were collected using Agilent Mass Hunter Workstation and analyzed using the software package Agilent Mass Hunter Quant B.07.00. Sphingolipids were quantitated based on 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), glucosyl ceramides (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, 24, and 24:1 carbon length were reported. Median (interquartile range) coefficient of variation (11.76, 6.85-20.53) are comparable with previously published sphingolipid data (66).

To calculate the Ceramide Risk Score (CERT1) that is in clinical use (50), we calculated C16:0, C18:0 and C24:1 concentration and their ratio to C24:0, assigning 2 points to those with levels in the 4^{th} quartile, 1 point to the 3^{rd} quartile, and 0 points to the bottom two quartiles, with total CERT1 scores ranging from 0-12. (22).

Statistics

Participant characteristics were summarized as mean ± standard deviation for continuous variables or N (%) for categorical variables (*Table 1*). Differences between cases and controls were compared using the Student *t*-test (two tailed) for continuous variables and chi-square test for categorical variables. P-values > 0.05 were considered significant. Lipid

species were summarized as medians and interquartile ranges (IQR) 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, variables were summed preceding log transformation.

Multivariable-adjusted and unadjusted odds ratios (ORs) and 95% confidence intervals (CI) were estimated using logistic regression and reported per standard deviation (of lipid species). *A priori*-defined covariates based on current American College of Cardiology (ACC) and American Heart Association (AHA) guidelines were considered in stepwise variable selection modeling (*Supplemental Table 3)*. These covariates included the following: age, sex, BMI, total cholesterol (total-C), low density lipoprotein cholesterol (LDL-C), high density lipoprotein cholesterol (HDL-C), very low-density lipoprotein cholesterol (VLDL cholesterol), triglycerides, hypertension, diabetes, and smoking. We calculated the percent change in odds ratio 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 evaluating significance of the interaction term using a likelihood ratio test (*Supplemental Table 4*)(67). Where effect modification was present (p-value for the interaction term<0.05, *Supplemental Table 4A*), 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 subgroups of the effect modifier variable (*Supplemental Table 4B, 4C*).

We applied machine learning (24, 25) to identify the most predictive biomarkers. To compare classical variable reduction techniques to our machine learning approaches, we performed a stepwise (forwards and backwards) regression (*Supplemental Table 5*). We then performed Least Absolute Shrinkage and Selection Operator (LASSO) regression (68) *(Supplemental Table 6*) and Random Forest analysis (*Supplemental Table 7*) (69). AHA/ACC lipid risk factor variables (LDL-cholesterol, 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%) datasets. For LASSO, 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 percentage of correctly identified cases/controls, averaging the percentage across ten training and testing splits. There were two data input approaches for Random Forest analysis. For a sphingolipid-only input, 32 sphingolipid variables were utilized, 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, triglycerides, etc.). To examine conditional correlations ($r \geq 0.20$) between ceramides and conventional biomarkers in CAD cases, we generated a gaussian graphical model (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 performed GGM in order to see whether our sphingolipid panel was redundant in the presence of traditional clinical lipid biomarkers (i.e. whether they are highly correlated, conditioned on the presence of all other lipids).

To compare the ability of different clinical markers and scores to distinguish between true cases and controls, we employed Receiver Operating Characteristic (ROC) – Area under the Curve (AUC) analysis and calculated the Net Reclassification Index (NRI), and Integrated Discrimination Index (IDI) (*Supplemental Table 1*) (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 false discovery rate (FDR) <0.05 to control for multiple statistical tests (*Supplemental Table 8*).

Study Approval

Lipid quantification and secondary data analysis on these patient samples were approved by the Institutional Review Board at the University of Utah and all patients provided written informed consent.

AUTHOR CONTRIBUTIONS

A.P, B.J.H., M.C.P, W.L.H and S.A.S conceived the project, designed the experiments, processed the ceramide data, and wrote the manuscript. A.P, J.A.M, and J.E.C. conducted the lipidomic analyses. P.N.H. and S.C.H. developed the CAD cohort including recruitment of patients, collection of clinical and demographic data, and storing of samples.

ACKNOWLEDGEMENTS

We are grateful for the support from the Metabolomics Core at the Health Sciences Center of the University of Utah. Funding: The authors receive research support from the National Institutes of Health (DK112826 and DK108833 to WLH and DK115824, DK116888,

and DK116450 to SAS), the Juvenile Diabetes Research Foundation (JDRF 3-SRA-2019-768-A-B to WLH), the American Diabetes Association (to SAS), the American Heart Association (to SAS), the Margolis Foundation (to SAS), the National Cancer Institute (5R00CA218694-03 to MCP), and the Huntsman Cancer Institute Cancer Center Support Grant (P30CA040214 to MCP).

REFERENCES

1. Rautou P-E, Vion A-C, Amabile N, Chironi G, Simon A, Tedgui A, et al. Microparticles, vascular function, and atherothrombosis. 2011;109(5):593-606.
2. Bergman BC, Brozinick JT, Strauss A, Bacon S, Kerege A, Bui HH, et al. Serum sphingolipids: relationships to insulin sensitivity and changes with exercise in humans. *Am J Physiol Endocrinol Metab.* 2015;309(4):E398-408.
3. Association AH. *American Heart Association CVD Burden Report* 2017.
4. Summers SA. Could Ceramides Become the New Cholesterol? *Cell Metab.* 2018;27(2):276-80.
5. Summers SA, Chaurasia B, and Holland WL. Metabolic Messengers: Ceramides. *Nature Metabolism.* 2019;(in press).
6. Merrill AH, Jr. De novo sphingolipid biosynthesis: a necessary, but dangerous, pathway. *J Biol Chem.* 2002;277(29):25843-6.
7. Chaurasia B, Tippetts TS, Monibas RM, Liu J, Li Y, Wang L, et al. Targeting a ceramide double bond improves insulin resistance and hepatic steatosis. *Science.* 2019.
8. Zelnik ID, Rozman B, Rosenfeld-Gur E, Ben-Dor S, and Futerman AH. A Stroll Down the CerS Lane. *Adv Exp Med Biol.* 2019;1159:49-63.
9. Hammerschmidt P, Ostkotte D, Nolte H, Gerl MJ, Jais A, Brunner HL, et al. CerS6-Derived Sphingolipids Interact with Mff and Promote Mitochondrial Fragmentation in Obesity. *Cell.* 2019;177(6):1536-52 e23.
10. Raichur S, Brunner B, Bielohuby M, Hansen G, Pfenninger A, Wang B, 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, Hammerschmidt P, Chen W, Jais A, Timper K, Awazawa M, et al. CerS1-Derived C18:0 Ceramide in Skeletal Muscle Promotes Obesity-Induced Insulin Resistance. *Cell Rep.* 2019;26(1):1-10 e7.
12. Turpin SM, Nicholls HT, Willmes DM, Mourier A, Brodesser S, Wunderlich CM, et al. Obesity-induced CerS6-dependent C16:0 ceramide production promotes weight gain and glucose intolerance. *Cell metabolism.* 2014;20(4):678-86.
13. Hla T, and Kolesnick R. C16:0-ceramide signals insulin resistance. *Cell Metab.* 2014;20(5):703-5.

14. Raichur S, Wang ST, Chan PW, Li Y, Ching J, Chaurasia B, et al. CerS2 haploinsufficiency inhibits beta-oxidation and confers susceptibility to diet-induced steatohepatitis and insulin resistance. *Cell Metab.* 2014;20(4):687-95.
15. Chaurasia B, and Summers SA. Ceramides - Lipotoxic Inducers of Metabolic Disorders. *Trends Endocrinol Metab.* 2015;26(10):538-50.
16. de Carvalho LP, Tan SH, Ow GS, Tang Z, Ching J, Kovalik JP, et al. Plasma Ceramides as Prognostic Biomarkers and Their Arterial and Myocardial Tissue Correlates in Acute Myocardial Infarction. *JACC Basic to translational science.* 2018;3(2):163-75.
17. Hilvo M, Meikle PJ, Pedersen ER, Tell GS, Dhar I, Brenner H, et al. Development and validation of a ceramide- and phospholipid-based cardiovascular risk estimation score for coronary artery disease patients. *Eur Heart J.* 2019.
18. Laaksonen R, Ekroos K, Sysi-Aho M, Hilvo M, Vihervaara T, Kauhanen D, et al. Plasma ceramides predict cardiovascular death in patients with stable coronary artery disease and acute coronary syndromes beyond LDL-cholesterol. *European heart journal.* 2016;37(25):1967-76.
19. Peterson LR, Xanthakis V, Duncan MS, Gross S, Friedrich N, Volzke H, et al. Ceramide Remodeling and Risk of Cardiovascular Events and Mortality. *Journal of the American Heart Association.* 2018;7(10).
20. Havulinna AS, Sysi-Aho M, Hilvo M, Kauhanen D, Hurme R, Ekroos K, et al. Circulating Ceramides Predict Cardiovascular Outcomes in the Population-Based FINRISK 2002 Cohort. *Arterioscler Thromb Vasc Biol.* 2016;36(12):2424-30.
21. Kaasenbrood L, Boekholdt SM, van der Graaf Y, Ray KK, Peters RJ, Kastelein JJ, et al. Distribution of Estimated 10-Year Risk of Recurrent Vascular Events and Residual Risk in a Secondary Prevention Population. *Circulation.* 2016;134(19):1419-29.
22. Kauhanen D, Sysi-Aho M, Koistinen KM, Laaksonen R, Sinisalo J, and Ekroos K. Development and validation of a high-throughput LC-MS/MS assay for routine measurement of molecular ceramides. *Analytical and bioanalytical chemistry.* 2016;408(13):3475-83.
23. Gotto AM, Jr. Interrelationship of triglycerides with lipoproteins and high-density lipoproteins. *The American journal of cardiology.* 1990;66(6):20a-3a.
24. Bzdok D, Altman N, and Krzywinski M. Statistics versus machine learning. *Nature methods.* 2018;15(4):233-4.
25. Song X, Mitnitski A, Cox J, and Rockwood K. Comparison of machine learning techniques with classical statistical models in predicting health outcomes. *Studies in health technology and informatics.* 2004;107(Pt 1):736-40.
26. Pencina MJ1 DAR, Pencina KM, Janssens AC, Greenland P. Interpreting incremental value of markers added to risk prediction models. *Am J Epidemiol.* 2012.
27. Pencina MJ, D'Agostino RB, Pencina KM, Janssens AC, and Greenland P. Interpreting incremental value of markers added to risk prediction models. *Am J Epidemiol.* 2012;176(6):473-81.
28. Hall SL, and Lorenc T. Secondary Prevention of Coronary Artery Disease. *Am Fam Physician* 2010;81(3):289-96.

29. Anroedh S, Hilvo M, Akkerhuis KM, Kauhanen D, Koistinen K, Oemrawsingh R, et al. Plasma concentrations of molecular lipid species predict long-term clinical outcome in coronary artery disease patients. *J Lipid Res.* 2018;59(9):1729-37.
30. Mantovani A, Bonapace S, Lunardi G, Salgarello M, Dugo C, Gori S, 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):2854-61.
31. Hilvo M, Simolin H, Metso J, Ruuth M, Oorni K, Jauhiainen M, et al. PCSK9 inhibition alters the lipidome of plasma and lipoprotein fractions. *Atherosclerosis.* 2018;269:159-65.
32. Meeusen JW, Donato LJ, Bryant SC, Baudhuin LM, Berger PB, and Jaffe AS. Plasma Ceramides. *Arterioscler Thromb Vasc Biol.* 2018;38(8):1933-9.
33. Lemaitre RN, Yu C, Hoofnagle A, Hari N, Jensen PN, Fretts AM, et al. Circulating Sphingolipids, Insulin, HOMA-IR, and HOMA-B: The Strong Heart Family Study. *Diabetes.* 2018;67(8):1663-72.
34. Wigger L, Cruciani-Guglielmacci C, Nicolas A, Denom J, Fernandez N, Fumeron F, et al. Plasma Dihydroceramides Are Diabetes Susceptibility Biomarker Candidates in Mice and Humans. *Cell Rep.* 2017;18(9):2269-79.
35. Jensen PN, Fretts AM, Yu C, Hoofnagle AN, Umans JG, Howard BV, et al. Circulating sphingolipids, fasting glucose, and impaired fasting glucose: The Strong Heart Family Study. *EBioMedicine.* 2019;41:44-9.
36. Lee SY, Kim JR, Hu Y, Khan R, Kim SJ, Bharadwaj KG, et al. Cardiomyocyte specific deficiency of serine palmitoyltransferase subunit 2 reduces ceramide but leads to cardiac dysfunction. *J Biol Chem.* 2012;287(22):18429-39.
37. Park TS, and Goldberg IJ. Sphingolipids, lipotoxic cardiomyopathy, and cardiac failure. *Heart Fail Clin.* 2012;8(4):633-41.
38. Park TS, Hu Y, Noh HL, Drosatos K, Okajima K, Buchanan J, et al. Ceramide is a cardiotoxin in lipotoxic cardiomyopathy. *J Lipid Res.* 2008;49(10):2101-12.
39. Ussher JR, Koves TR, Cadete VJ, Zhang L, Jaswal JS, Swyrd SJ, et al. Inhibition of de novo ceramide synthesis reverses diet-induced insulin resistance and enhances whole-body oxygen consumption. *Diabetes.* 2010;59(10):2453-64.
40. Glaros EN, Kim WS, Quinn CM, Jessup W, Rye KA, and Garner B. Myriocin slows the progression of established atherosclerotic lesions in apolipoprotein E gene knockout mice. *J Lipid Res.* 2008;49(2):324-31.
41. Glaros EN, Kim WS, Wu BJ, Suarna C, Quinn CM, Rye KA, et al. Inhibition of atherosclerosis by the serine palmitoyl transferase inhibitor myriocin is associated with reduced plasma glycosphingolipid concentration. *Biochem Pharmacol.* 2007;73(9):1340-6.
42. Hojjati MR, Li Z, Zhou H, Tang S, Huan C, Ooi E, et al. Effect of myriocin on plasma sphingolipid metabolism and atherosclerosis in apoE-deficient mice. *J Biol Chem.* 2005;280(11):10284-9.
43. Park TS, Rosebury W, Kindt EK, Kowala MC, and 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, Brozinick JT, Wang LP, Hawkins ED, Sargent KM, Liu Y, et al. Inhibition of ceramide synthesis ameliorates glucocorticoid-, saturated-fat-, and obesity-induced insulin resistance. *Cell Metab.* 2007;5(3):167-79.
45. Park TS, Panek RL, Rekhter MD, Mueller SB, Rosebury WS, Robertson A, et al. Modulation of lipoprotein metabolism by inhibition of sphingomyelin synthesis in ApoE knockout mice. *Atherosclerosis.* 2006;189(2):264-72.
46. Bharath LP, Ruan T, Li Y, Ravindran A, Wan X, Nhan JK, et al. Ceramide-Initiated Protein Phosphatase 2A Activation Contributes to Arterial Dysfunction In Vivo. *Diabetes.* 2015;64(11):3914-26.
47. Zhang QJ, Holland WL, Wilson L, Tanner JM, Kearns D, Cahoon JM, et al. Ceramide mediates vascular dysfunction in diet-induced obesity by PP2A-mediated dephosphorylation of the eNOS-Akt complex. *Diabetes.* 2012;61(7):1848-59.
48. Bikman BT, Guan Y, Shui G, Siddique MM, Holland WL, Kim JY, et al. Fenretinide prevents lipid-induced insulin resistance by blocking ceramide biosynthesis. *J Biol Chem.* 2012;287(21):17426-37.
49. Holland WL, Bikman BT, Wang LP, Yuguang G, Sargent KM, Bulchand S, 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):1858-70.
50. Vasiliauskaite-Brooks I, Sounier R, Rochaix P, Bellot G, Fortier M, Hoh F, et al. Structural insights into adiponectin receptors suggest ceramidase activity. *Nature.* 2017;544(7648):120-3.
51. Holland WL, Miller RA, Wang ZV, Sun K, Barth BM, Bui HH, et al. Receptor-mediated activation of ceramidase activity initiates the pleiotropic actions of adiponectin. *Nature medicine.* 2011;17(1):55-63.
52. Holland WL, Adams AC, Brozinick JT, Bui HH, Miyauchi Y, Kusminski CM, et al. An FGF21-adiponectin-ceramide axis controls energy expenditure and insulin action in mice. *Cell Metab.* 2013;17(5):790-7.
53. Weng SF, Reps J, Kai J, Garibaldi JM, and Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PloS one.* 2017;12(4):e0174944.
54. Wagner-Golbs A, Neuber S, Kamlage B, Christiansen N, Bethan B, Rennefahrt U, et al. Effects of Long-Term Storage at -80 degrees C on the Human Plasma Metabolome. *Metabolites.* 2019;9(5).
55. Brunkhorst R, Pfeilschifter W, Patyna S, Buttner S, Eckes T, Trautmann S, et al. Preanalytical Biases in the Measurement of Human Blood Sphingolipids. *International journal of molecular sciences.* 2018;19(5).
56. Aronson JK, and Ferner RE. Biomarkers-A General Review. *Current protocols in pharmacology.* 2017;76:9.23.1-9..17.
57. Hopkins PN, Wu LL, Hunt SC, and Brinton EA. Plasma triglycerides and type III hyperlipidemia are independently associated with premature familial coronary artery disease. *Journal of the American College of Cardiology.* 2005;45(7):1003-12.
58. Hunt SC, Williams RR, and Barlow GK. A comparison of positive family history definitions for defining risk of future disease. *Journal of chronic diseases.* 1986;39(10):809-21.

59. Vikulova DN, Grubisic M, Zhao Y, Lynch K, Humphries KH, Pimstone SN, et al. Premature Atherosclerotic Cardiovascular Disease: Trends in Incidence, Risk Factors, and Sex-Related Differences, 2000 to 2016. *Journal of the American Heart Association.* 2019;8(14):e012178.
60. Higgins M, Province M, Heiss G, Eckfeldt J, Ellison RC, Folsom AR, et al. NHLBI Family Heart Study: objectives and design. *American journal of epidemiology.* 1996;143(12):1219-28.
61. Hopkins PN, Nanjee MN, Wu LL, McGinty MG, Brinton EA, Hunt SC, et al. Altered composition of triglyceride-rich lipoproteins and coronary artery disease in a large case-control study. *Atherosclerosis.* 2009;207(2):559-66.
62. Clinics LR. DHEW publication NIH; 1974:75-628.
63. Warnick GR, Benderson J, and Albers JJ. Dextran sulfate-Mg2+ precipitation procedure for quantitation of high-density-lipoprotein cholesterol. *Clinical chemistry.* 1982;28(6):1379-88.
64. Wu LL, Warnick GR, Wu JT, Williams RR, and Lalouel JM. A rapid micro-scale procedure for determination of the total lipid profile. *Clinical chemistry.* 1989;35(7):1486-91.
65. Stratford S, Hoehn KL, Liu F, and 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):36608-15.
66. Sullards MC, Liu Y, Chen Y, and Merrill AH, Jr. 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):838-53.
67. Achim Zeilies TH. *R News*. 2002:7-10.
68. Friedman J HT, Tibshirani R Regularization Paths for Generalized Linear Models via Coordinate Descent *Journal of Statistical Software* 2010;1(33):1-22.
69. Liaw A WM. Classification and Regression by randomForest. *R News.* 2002;2(3):18-22.
70. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research.* 2003;13(11):2498-504.
71. C B. _network: Classes for Relational Data_. *The Statnet Project.* 2015.
72. C B. network: a Package for Managing Relational Data in R. *Journal of Statistical Software.* 2008;24(2).
73. Genz A, and Bretz F. Computation of Multivariate Normal and t Probabilities *Lecture Notes in Statistics.* 2009;195.
74. Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, et al. mvtnorm: Multivariable Normal and t Distributions. *R package version10-11.* 2019.
75. Tibshirani JFaTHaR. glasso: Graphical Lasso: Estimation of Gaussian Graphical Models. 2018.
76. Krumsiek J SK, Illig T, Adamski J, Theis FJ. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC Syst Biol* 2011.
77. Robin X TN, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* 2011;12:77.
78. R Development Core Team. R Foundation for Statistical Computing 2018.

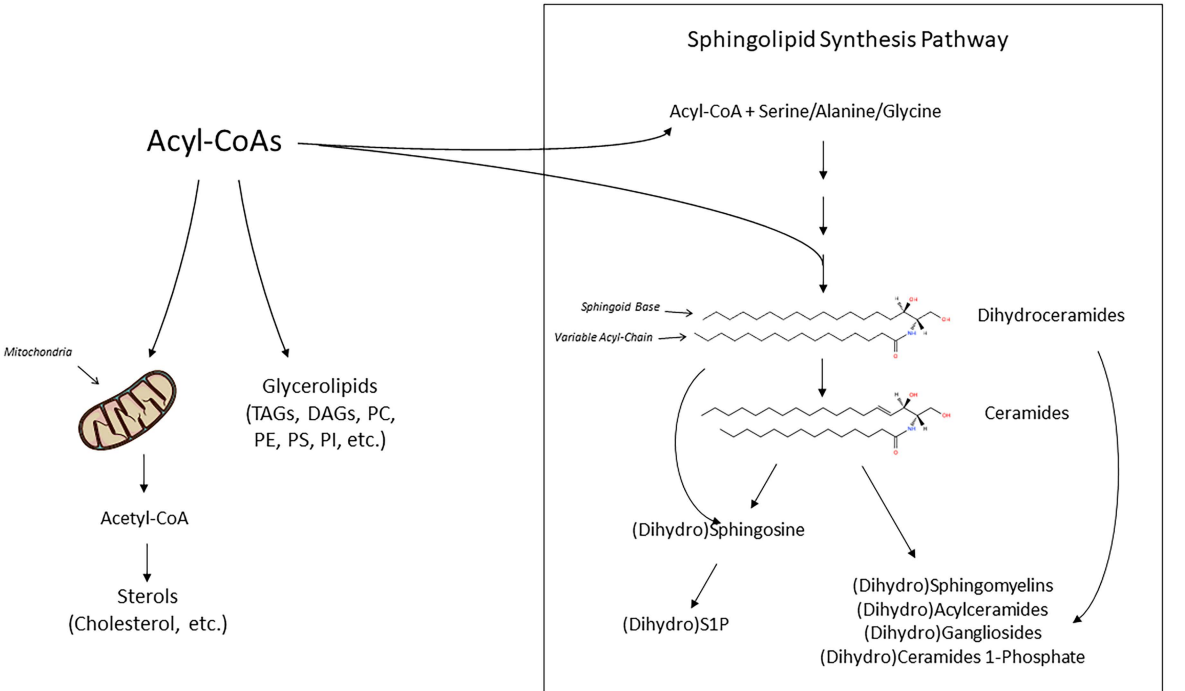

Acyl-CoAs
Mitochondria
Glycerolipids
(TAGs, DAGs, PC,
PE, PS, PI, etc.)
Acetyl-CoA
Sterols
(Cholesterol, etc.)
Sphingolipid Synthesis Pathway
Acyl-CoA + Serine/Alanine/Glycine
Sphingoid Base
Variable Acyl-Chain
Dihydroceramides
Ceramides
(Dihydro)Sphingosine
(Dihydro)S1P
(Dihydro)Sphingomyelins
(Dihydro)Acylceramides
(Dihydro)Gangliosides
(Dihydro)Ceramides 1-Phosphate

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 lipid that represent a minor subset of the lipidome but play critical roles in signaling events.

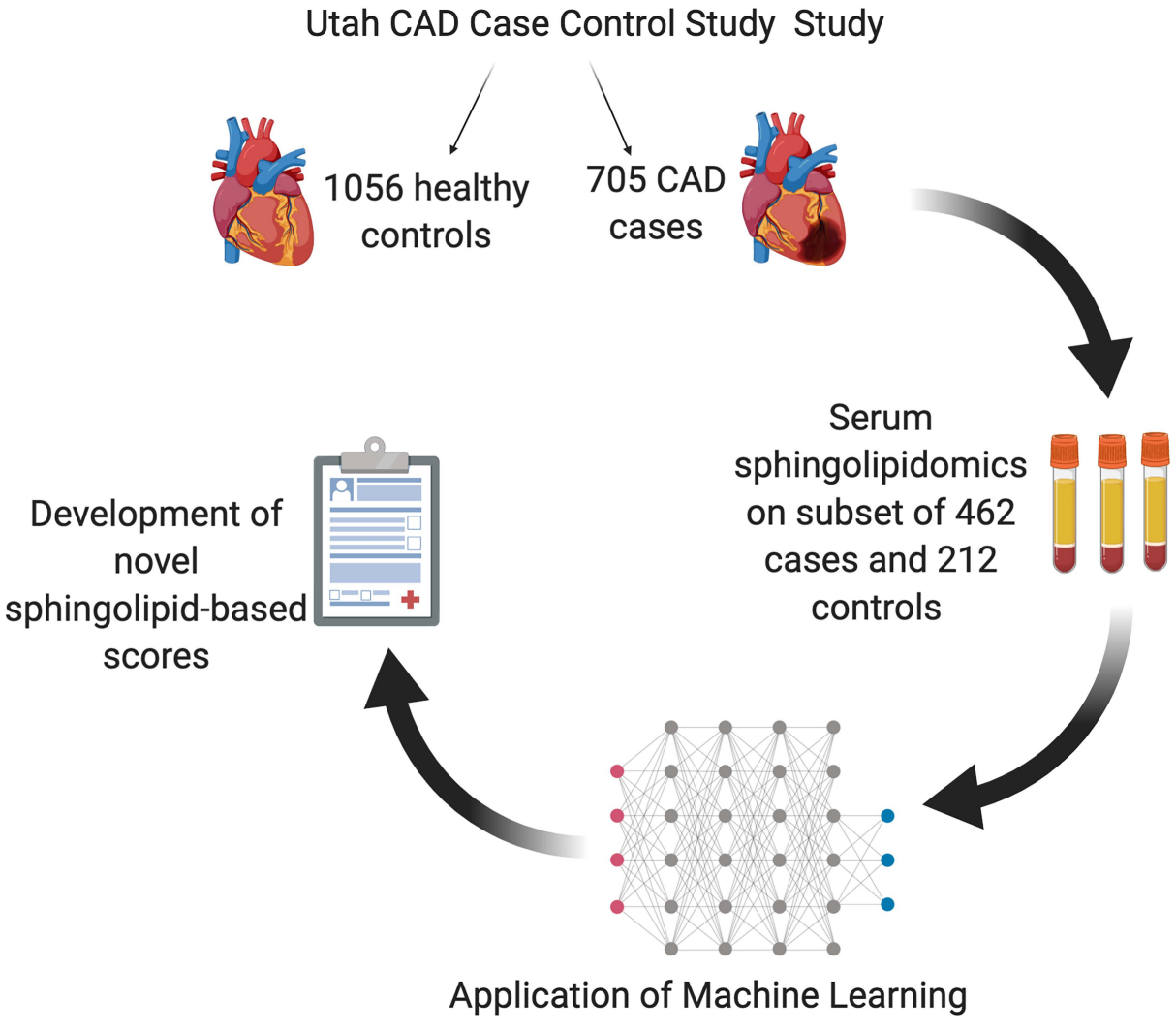

Utah CAD Case Control Study Study
1056 healthy controls
705 CAD cases
Serum sphingolipidomics on subset of 462 cases and 212 controls
Application of Machine Learning
Development of novel sphingolipid-based scores

Figure 2. Schematic of Utah coronary artery disease (CAD) study design and the subset of available biospecimens used for liquid chromatography tandem mass spectrometry 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).

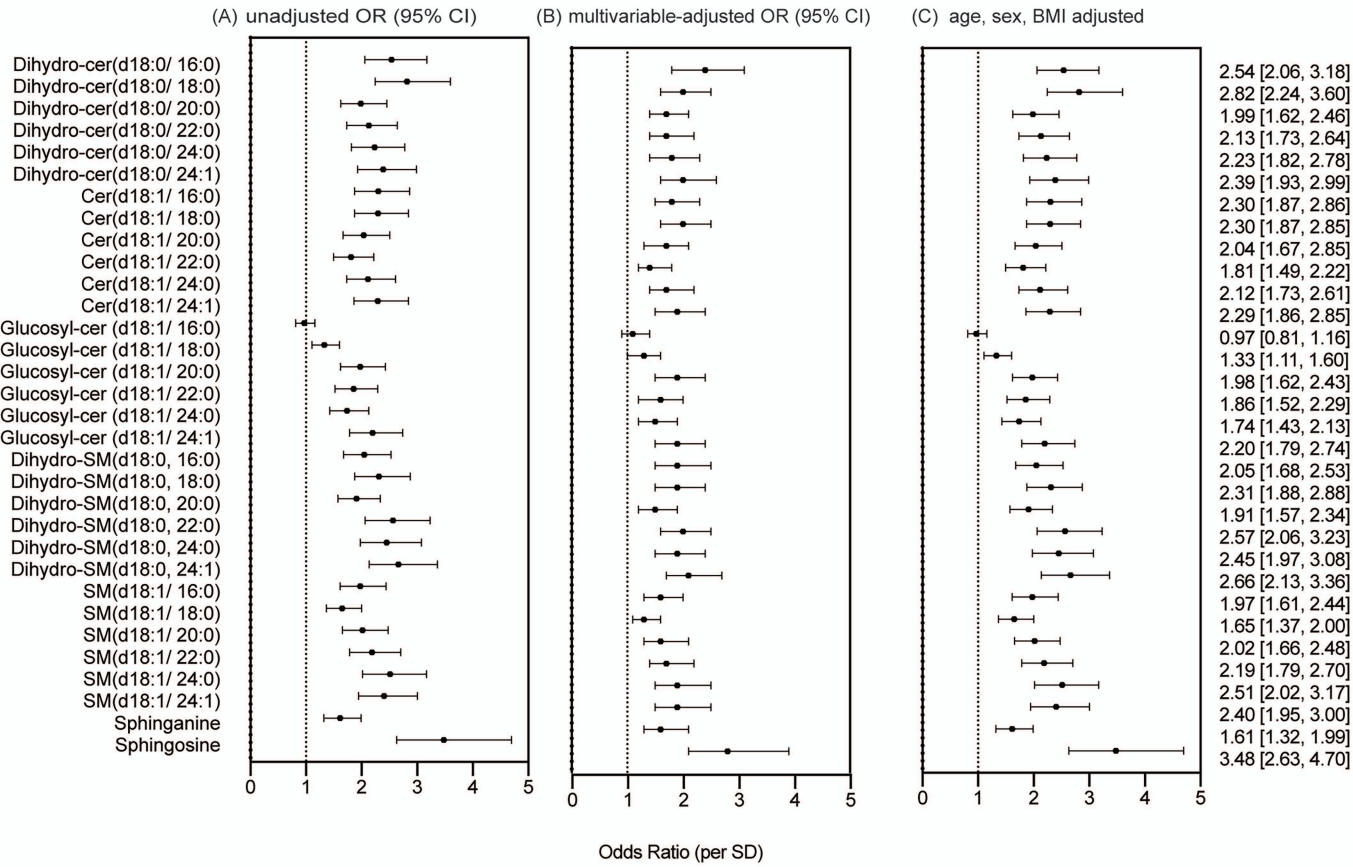

(A) unadjusted OR (95% CI)
(B) multivariable-adjusted OR (95% CI)
(C) age, sex, BMI adjusted
Dihydro-cer(d18:0/ 16:0)
Dihydro-cer(d18:0/ 18:0)
Dihydro-cer(d18:0/ 20:0)
Dihydro-cer(d18:0/ 22:0)
Dihydro-cer(d18:0/ 24:0)
Dihydro-cer(d18:0/ 24:1)
Cer(d18:1/ 16:0)
Cer(d18:1/ 18:0)
Cer(d18:1/ 20:0)
Cer(d18:1/ 22:0)
Cer(d18:1/ 24:0)
Cer(d18:1/ 24:1)
Glucosyl-cer (d18:1/ 16:0)
Glucosyl-cer (d18:1/ 18:0)
Glucosyl-cer (d18:1/ 20:0)
Glucosyl-cer (d18:1/ 22:0)
Glucosyl-cer (d18:1/ 24:0)
Glucosyl-cer (d18:1/ 24:1)
Dihydro-SM(d18:0, 16:0)
Dihydro-SM(d18:0, 18:0)
Dihydro-SM(d18:0, 20:0)
Dihydro-SM(d18:0, 22:0)
Dihydro-SM(d18:0, 24:0)
Dihydro-SM(d18:0, 24:1)
SM(d18:1/ 16:0)
SM(d18:1/ 18:0)
SM(d18:1/ 20:0)
SM(d18:1/ 22:0)
SM(d18:1/ 24:0)
SM(d18:1/ 24:1)
Sphinganine
Sphingosine
2.54 [2.06, 3.18]
2.82 [2.24, 3.60]
1.99 [1.62, 2.46]
2.13 [1.73, 2.64]
2.23 [1.82, 2.78]
2.39 [1.93, 2.99]
2.30 [1.87, 2.86]
2.30 [1.87, 2.85]
2.04 [1.67, 2.85]
1.81 [1.49, 2.22]
2.12 [1.73, 2.61]
2.29 [1.86, 2.85]
0.97 [0.81, 1.16]
1.33 [1.11, 1.60]
1.98 [1.62, 2.43]
1.86 [1.52, 2.29]
1.74 [1.43, 2.13]
2.20 [1.79, 2.74]
2.05 [1.68, 2.53]
2.31 [1.88, 2.88]
1.91 [1.57, 2.34]
2.57 [2.06, 3.23]
2.45 [1.97, 3.08]
2.66 [2.13, 3.36]
1.97 [1.61, 2.44]
1.65 [1.37, 2.00]
2.02 [1.66, 2.48]
2.19 [1.79, 2.70]
2.51 [2.02, 3.17]
2.40 [1.95, 3.00]
1.61 [1.32, 1.99]
3.48 [2.63, 4.70]
0
1
2
3
4
5
Odds Ratio (per SD)

Figure 3. Forest plot of OR (95% CI) for coronary artery disease (CAD) per standard deviation of sphingolipid species in the Utah CAD study (A) unadjusted; (B) fully-adjusted (age, sex, BMI, total-C, LDL-C, VLDL-C, triglycerides, hypertension, diabetes, smoking); (C) minimally-adjusted (age, sex, BMI) model. OR (95% CI) numerically presented represent the minimally-adjusted age, sex, BMI model. Dihydro cer, dihydro ceramide; cer, ceramide

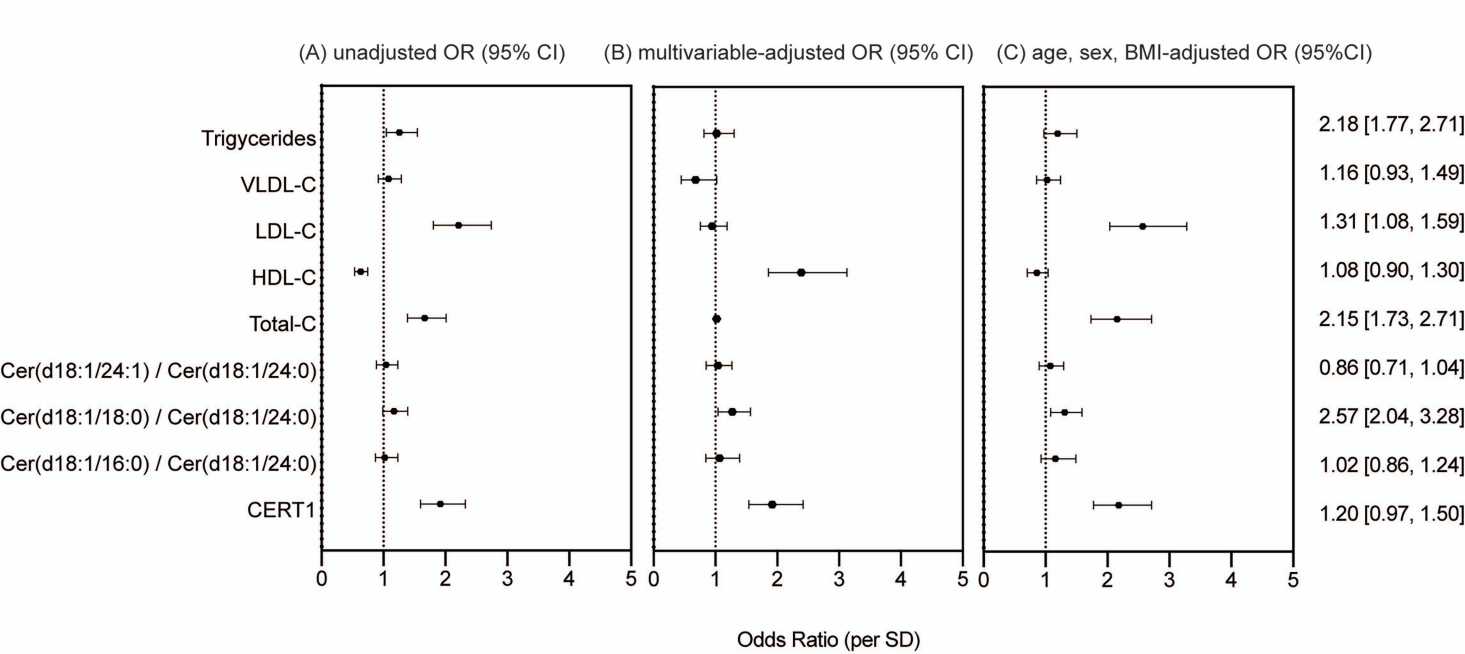

(A) unadjusted OR (95% CI)
(B) multivariable-adjusted OR (95% CI)
(C) age, sex, BMI-adjusted OR (95%CI)
Trigycerides
VLDL-C
LDL-C
HDL-C
Total-C
Cer(d18:1/24:1) / Cer(d18:1/24:0)
Cer(d18:1/18:0) / Cer(d18:1/24:0)
Cer(d18:1/16:0) / Cer(d18:1/24:0)
CERT1
2.18 [1.77, 2.71]
1.16 [0.93, 1.49]
1.31 [1.08, 1.59]
1.08 [0.90, 1.30]
2.15 [1.73, 2.71]
0.86 [0.71, 1.04]
2.57 [2.04, 3.28]
1.02 [0.86, 1.24]
1.20 [0.97, 1.50]
0
1
2
3
4
5
Odds Ratio (per SD)

Figure 4. OR (95% CI) of CAD per standard deviation of previously reported lipid markers of cardiovascular risk in the Utah CAD study (A) unadjusted; (B) fully-adjusted (age, sex, BMI, hypertension, diabetes, smoking); and (C) minimally-adjusted (age, sex, BMI). OR(95% CI) numerically presented represent the minimally-adjusted age, sex, BMI model. Since we compared clinical lipid markers (LDL, VLDL, HDL, triglycerides) to ceramide ratios and scores they were not included in the fully-adjusted model. CERT1, cardiac event risk test (12-point scale); HDL cholesterol, high density lipoprotein cholesterol (mg/dL); LDL cholesterol, low density lipoprotein cholesterol (mg/dL); VLDL cholesterol, very low density lipoprotein cholesterol (mg/dL), triglycerides (mg/dL)

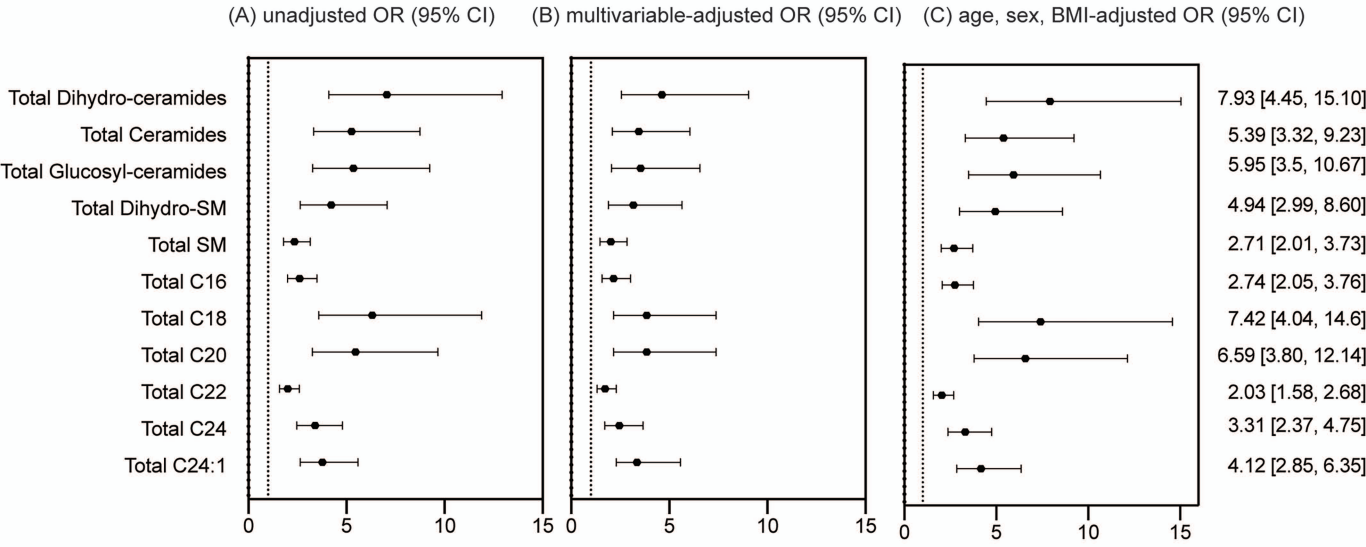

(A) unadjusted OR (95% CI)
(B) multivariable-adjusted OR (95% CI)
(C) age, sex, BMI-adjusted OR (95% CI)
Total Dihydro-ceramides
Total Ceramides
Total Glucosyl-ceramides
Total Dihydro-SM
Total SM
Total C16
Total C18
Total C20
Total C22
Total C24
Total C24:1
7.93 [4.45, 15.10]
5.39 [3.32, 9.23]
5.95 [3.5, 10.67]
4.94 [2.99, 8.60]
2.71 [2.01, 3.73]
2.74 [2.05, 3.76]
7.42 [4.04, 14.6]
6.59 [3.80, 12.14]
2.03 [1.58, 2.68]
3.31 [2.37, 4.75]
4.12 [2.85, 6.35]
0
5
10
15

Figure 5. OR (95% CI) of CAD per standard deviation of summed sphingolipid variables in the Utah CAD study (A) unadjust-ed; (B) fully-adjusted (age, sex, BMI, total-C, LDL-C, VLDL-C, triglycerides, hypertension, diabetes, smoking); (C) minimally-adjusted (age, sex, BMI). OR(95% CI) numerically presented represent the minimally-adjusted age, sex, 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.

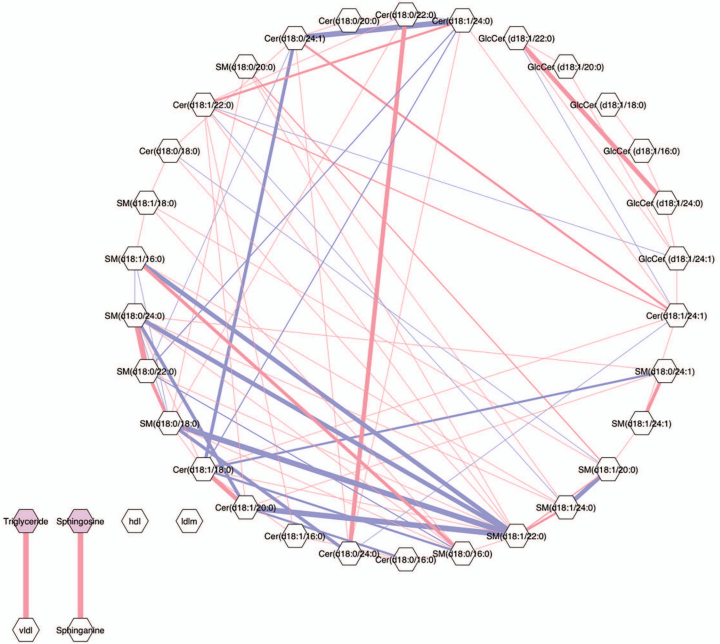

Cer(d18:0/22:0)
Cer(d18:1/24:0)
Cer(d18:0/20:0)
Cer(d18:0/24:1)
GlcCer(d18:1/22:0)
SM(d18:0/20:0)
GlcCer(d18:1/20:0)
Cer(d18:1/22:0)
GlcCer(d18:1/18:0)
Cer(d18:0/18:0)
GlcCer(d18:1/16:0)
SM(d18:1/18:0)
GlcCer(d18:1/24:0)
SM(d18:1/16:0)
GlcCer(d18:1/24:1)
SM(d18:0/24:0)
Cer(d18:1/24:1)
SM(d18:0/22:0)
SM(d18:0/24:1)
SM(d18:0/18:0)
SM(d18:1/24:1)
Cer(d18:1/18:0)
SM(d18:1/20:0)
Cer(d18:1/20:0)
SM(d18:1/24:0)
Cer(d18:1/16:0)
SM(d18:1/22:0)
Cer(d18:0/24:0)
SM(d18:0/16:0)
Cer(d18:0/16:0)
Triglyceride
Sphingosine
hdl
ldlm
vldl
Sphinganine

Figure 6. Gaussian graphical model (GGM) of correlations between ceramide species and conventional lipid markers in Coronary Artery Disease (CAD) cases, conditioned on the presence of all other analytes ($r \geq 0.20$). Analytes are represent-ed by nodes (grey circles) and conditional correlations by edges (lines). Pink lines indicate positive correlations, blue inverse. Line width represents the strength of conditional correlation and the lack of a line indicates no detectable relationship above the threshold.

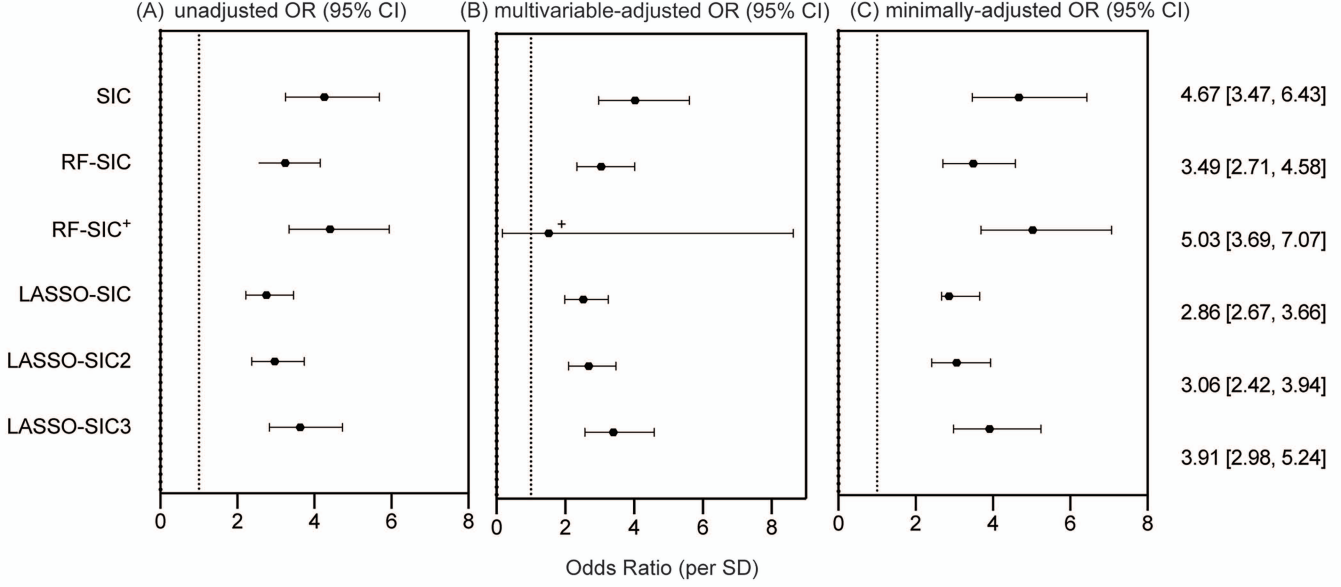
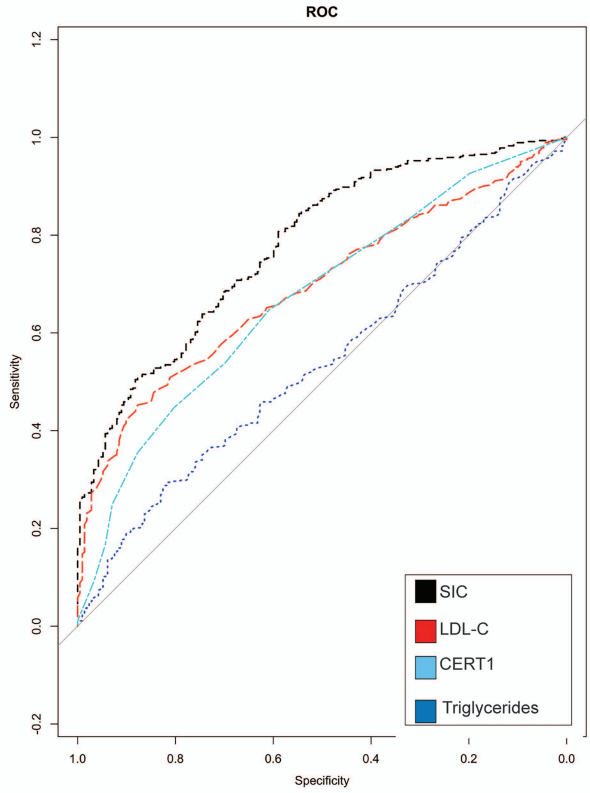
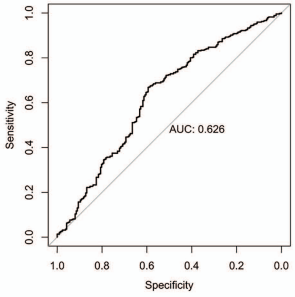

(A) unadjusted OR (95% CI)
(B) multivariable-adjusted OR (95% CI)
(C) minimally-adjusted OR (95% CI)
SIC
RF-SIC
RF-SIC+
LASSO-SIC
LASSO-SIC2
LASSO-SIC3
4.67 [3.47, 6.43]
3.49 [2.71, 4.58]
5.03 [3.69, 7.07]
2.86 [2.67, 3.66]
3.06 [2.42, 3.94]
3.91 [2.98, 5.24]
0
2
4
6
8
Odds Ratio (per SD)

Figure 7. OR (95% CI) of coronary artery disease (CAD) per standard deviation of novel scores generated through the application of machine learning approaches in the Utah CAD study (A) unadjusted; (B) multivariable-adjusted (age, sex, BMI, diabetes, hypertension, smoking); (C) minimally-adjusted (age, sex, BMI). The multivariable models for this analysis do not include HDL-C, LDL-C, VLDL-C, total-C, or triglycerides as they were included as input variables.

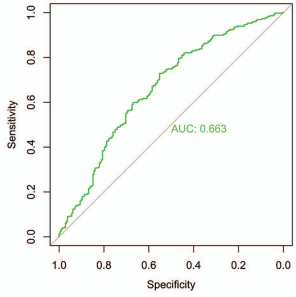
(A)



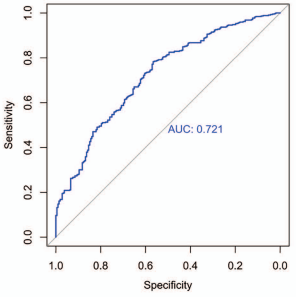
(C) AHA/ACC Clinical Characteristics



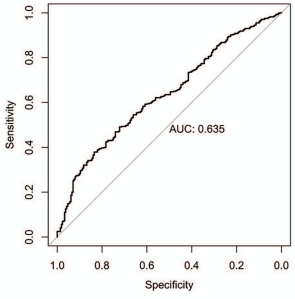
(D) AHA/ACC Clinical Characteristics + CERT1



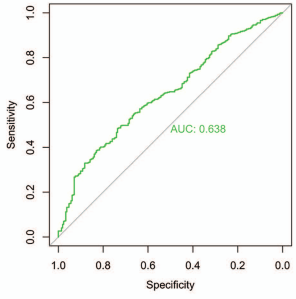
(E) AHA/ACC Clinical Characteristics + SIC



(F) AHA/ACC Clinical Characteristics/Lipids



(G) AHA/ACC Clinical Characteristics/Lipids + CERT1



(H) AHA/ACC Clinical Characteristics/Lipids + SIC

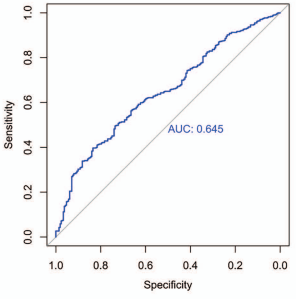


Figure 8. Comparing conventional coronary artery disease (CAD) risk markers to novel sphingolipid scores in the Utah CAD study. (A) Receiver operator characteristic (ROC) curve for novel sphingolipid inclusive CAD (SIC) score and conventional risk markers. (B) C-statis-tics for both conventional and novel risk markers for CAD. (C) ROC curve for American Heart Association (AHA)/American College of Cardiology (ACC) based clinical risk factors (age, sex, BMI, diabetes, hypertension, smoking) and (D) these same guidelines in addition to the cardiac event risk test (CERT1) score, and (E) the Sphingolipid Inclusive CAD (SIC) score. ROC curves for (F) the aforementioned AHA/ACC clinical markers in addition to lipid markers (total-C, HDL-C, LDL-C, VLDL-C, triglycerides), (G) the clinical and lipid markers in addition to CERT1, and (H) SIC. For (C)-(G) c-statistics are indicated on the respective graphs by area under the curve (AUC).

(A) minimally-adjusted OR (95% CI)

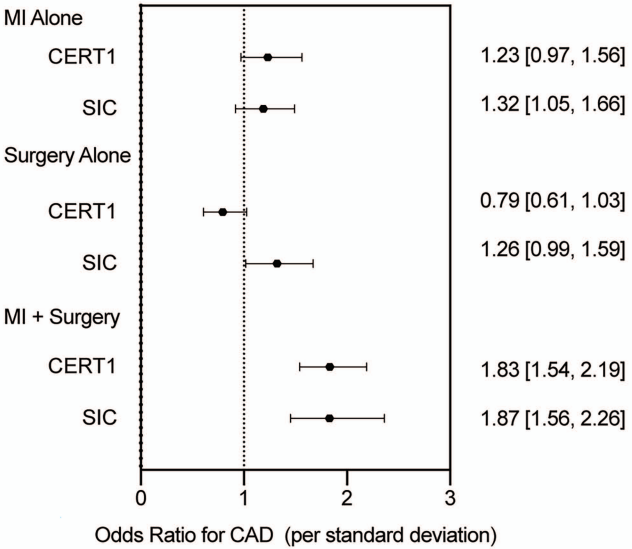


Figure 9. Association of sphingolipid scores with coronary artery disease (CAD), stratified by disease presentation (MI alone, surgery alone, MI plus surgery). (A) OR (95% CI) for (CAD) per standard deviation of sphingolipid species in the Utah CAD study adjusted for age, sex, BMI). MI, myocardial infarction; CERT1, ceramide risk score; SIC, sphingolipid inclusive CAD risk score.

Table 1. Baseline characteristics of case and control participants in the Utah coronary artery disease (CAD) study.

	Control	Case	P-Value
No. of Subjects	212	462	
Gender			
Male, *n* (%)	91 (43%)	356 (77%)	
Age (years) [A]	53.5 ± 6.9	55.6 ± 7.5	0.004
BMI [A]	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 Cholesterol [A]	189.5 ± 3.3	209.2 ± 4.6	<0.001
HDL	46.7 ± 12.8	40.9 ± 1.2	<0.001
LDL	103.1 ± 2.8	128.7 ± 4.2	<0.001
VLDL	37.3 ± 22.5	39.3 ± 3	0.326
Serum Triglycerides [A]	178 ± 9.5	202.7 ± 1.4	0.008

Clinical characteristics of Utah CAD study cases (*n* = 212) and controls (*n* = 462). Variables were compared between cases and controls using a two-tailed t-test for continuous variables and chi-square test for categorical variables with a p-value of <0.05 considered significant.

A mean ± standard deviation.

Age, years; BMI, body mass index; HDL high density lipoprotein (mg/dL); LDL, low density lipoprotein (md/dL); VLDL, very low-density lipoprotein (mg/dL); total cholesterol (mg/dL); serum triglycerides (mg/dL); NA, data not available

Table 2. Means and interquartile ranges for LC-MS/MS measured sphingolipids in and control groups of the Utah coronary artery disease (CAD) study.

Lipid	Control	Case	P-value
Dihydro Cer d18:0,16:0	0.1 (0.8-0.2)	0.2 (0.2-0.3)	$<2_E$-16
Dihydro Cer d18:0,18:0	0.09 (0.06-0.1)	0.1 (0.09-0.2)	$<2_E$-16
Dihydro Cer d18:0,20:0	0.05 (0.03-0.07)	0.07 (0.05-0.1)	2.36_E-10
Dihydro Cer d18:0,22:0	0.2 (0.1-0.3)	0.3 (0.2-0.5)	1.39_E-12
Dihydro Cer d18:0,24:0	0.4 (0.3-0.7)	0.7 (0.4-1.2)	1.29_E-14
Dihydro Cer d18:0,24:1	0.2 (0.1-0.4)	0.4 (0.2-0.7)	4.26_E-15
Cer d18:1,16:0	131.3 (87.8-201.3)	217.2 (150.3-324.9)	5.83_E-16
Cer d18:1,18:0	48 (29.9-76.7)	86 (53.9-138.3)	5.40_E-16
Cer d18:1,20:0	44.1 (30.6-65.6)	68.1 (43.8-102.3)	6.12_E-13
Cer d18:1,22:0	264.4 (185.9-436.7)	399.1 (278.8-631.3)	3.63_E-11
Cer d18:1,24:0	98 (64.5-148.3)	157.6 (106.8-245.1)	1.61_E-15
Cer d18:1,24:1	264.7 (173.5-411.8)	437.6 (306.2-669.2)	$<2_E$-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_E-08
GC Cer d18:1,22:0	494.7 (367.7-705)	713.5 (446.5-1022.2)	2.87_E-08
GC Cer d18:1,24:0	456.9 (321.1-591.1)	585.5 (408.2-879.4)	8.14_E-08
GC Cer d18:1 24:1	397.5 (287.4-547.1)	575.7 (399.9-866.1)	5.11_E-10
Dihydro SM d18:0,16:0	45 (30.9-64.1)	61.1 (45.9-88.9)	1.22_E-09
Dihydro SM d18:0,18:0	14 (7.5-25.2)	26.37 (14.3-49.6)	2.72_E-09
Dihydro SM d18:0,20:0	28.3 (12.9-48.2)	48.4 (25-91.2)	1.25_E-07
Dihydro SM d18:0,22:0	4.8 (2.2-10.2)	10.2 (5.9-21.4)	9.08_E-10
Dihydro SM d18:0,24:0	0.9 (0.4-1.5)	1.5 (0.9-2.8)	1.57_E-10
Dihydro SM d18:0,24:1	21.7 (10.7-42.1)	47.6 (26.3-92.2)	1.40_E-10
SM d18:1,16:0	592 (437.7-818.5)	779.3 (595-1066.3)	3.77_E-10
SM d18:1,18:0	155.1 (108.3-217.6)	186.7 (136.4-268.6)	2.54_E-06
SM d18:1,20:0	183 (78.7-354.3)	354.7 (183.3-698.6)	1.03_E-07
SM d18:1,22:0	359.4 (150.9-740)	718.7 (365.1-1462.2)	4.28_E-09
SM d18:1,24:0	154.2 (70.4-283.4)	293.8 (174.2-640.8)	1.44_E-09
SM d18:1,24:1	432.8 (200.5-879.1)	913.1 (478.2-1907.4)	1.85_E-09
Sphinganine	0.03 (0.02-0.04)	0.04 (0.03-0.06)	4.92_E-06
Sphingosine	0.08 (0.05-0.1)	0.15 (0.1-0.3)	$<2_E$-16

Two tailed t-test was used to compare case (n = 212) and control (n = 462) concentrations of LC-MS/MS measured sphingolipids. P-values are for the parsimonious age, sex, BMI adjusted model and are considered significant at a false discovery rate (FDR) <0.05. Lipid concentrations are represented here as mean (interquartile range). The fully adjusted model (i.e. age, sex, BMI, total-C, LDL-C, VLDL-C, triglycerides, hypertension, diabetes, smoking) were also run, but results were not materially different than the parsimonious model, so here we present only the minimally adjusted model.
BMI, body mass index; total-C, total cholesterol; LDL-C, low-density lipoprotein cholesterol; VLDL-C, very low-density lipoprotein cholesterol; Dihydro Cer, dihydro ceramide; Cer, ceramide; GC Cer, glucosyl ceramide; Dihydro SM, dihydro sphingomyelin; SM, sphingomyelin; units are pmol lipd/ml serum

Table 3. Novel Sphingolipid scores for coronary artery disease (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

Random forest (RF) and least absolute shrinkage and selection operator (LASSO) regression were applied for variable reduction.

Table 4. Area under the curve of receiver operator characteristic plots for lipid-based clinical indices.

Clinical Index	C-Statistic
Triglycerides	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.
SIC, sphingolipid inclusive CAD score

Table 5. Stratification of Utah coronary artery disease (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 ± .24	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 ± 402
Total-C	189.9 ± 3.3	192 ± 43.4	220.6 ± 49.1	210.2 ± 44.4
Triglycerides	178 ± 9.5	193 ± 133.2	175.8 ± 89.3	212.2 ± 152.3

Clinical lipid marker serum concentrations stratified by disease severity and presented as concentration ± standard deviation.
MI, myocardial infarction; Surgery, percutaneous transluminal coronary artery bypass grafting (PCTA) or coronary artery bypass grafting (CABG)