Supplemental Table 1. Characteristics of the study population of the Utah CAD study, (A) the total case-control study (Total CAD); (B) For participants with available biospecimens used for the current analysis (MS-CAD). Variables were compared between the two control groups using t-tests for continuous variables and chi-square tests for categorical variables. (B) describes data missing from the sample used for the current analysis. We conducted multiple imputation of these variables but it did not change the results.

(A)

Variable	Total CAD	MS-CAD	P-value
No. of Subjects	1057	212	
Gender			0.19
Male, *n* (%)	506 (48%)	91 (43%)	
Smoking			0.41
Yes, *n* (%)	242 (23%)	43 (20%)	
No, *n* (%)	815 (77%)	169 (80%)	
Diabetes			0.61
Yes, *n* (%)	62 (6%)	11 (5%)	
No, *n* (%)	995 (94%)	201 (95%)	
Hypertension			0.21
Yes, *n* (%)	266 (25%)	54 (26%)	
No, *n* (%)	791 (75%)	158 (74%)	
Lipid Lowering Medication			0.13
Yes, *n* (%)	99 (9%)	13 (6%)	
No, *n* (%)	958 (91%)	199 (94%)	
BMI	27.81	28.27	0.27
Total Cholesterol	187.79	189.53	0.49
Total Triglycerides	151.70	177.96	0.003

(B)

Variable	n (%) Missing
height	15 (2.2%)
weight	16 (2.4%)
BMI	16 (2.4%)
glucose	3 (0.4%)

Supplemental Table 2. To test the effect of including conventional lipid markers in the primary logistic regression model, we calculated ORs for sphingolipid associations with CAD using multivariable logistic regression. The following variables were included in the model, one at a time: (age (years), sex (male or female), body mass index (BMI, Kg/m2), smoking ("ever" or "never" to smoking daily for a year or more), 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 >100 mg/dL), and hypertension (prior physician diagnosis or blood pressure >140/90 mm Hg)). If inclusion of the variable in the model changed the OR >10% and/or the variable significantly improved model fit according to likelihood ratio test, it was retained. (A) The top row denotes the covariates included in the model, with + indicating an addition to age, sex, BMI. The values are percent changes of ORs from age, sex, BMI adjusted regression model. (B) Due to changes in OR >10%, we show the individual ORs (95% CI) with lower confidence interval (LCI) and upper confidence interval (UCI) for smoking, diabetes mellitus (DM), hypertension, LDL-C, total-C, and all of the variables.

(A) Percent change in risk in comparison to the age, sex, BMI adjusted model

	Unadjusted	+ HDL-C	+ LDL-C	+ VLDL-C	+ triglycerides	+ hypertension	+ diabetes mellitus	+ smoking	+all
Dihydro-cer(d18:0/ 16:0)	1.4	0.6	5.2	-4.5	-0.4	6.5	7.9	8.3	-18.0
Dihydro-cer(d18:0/ 18:0)	3.9	0.6	6.2	-4.5	-1.4	-3.7	-2.3	-5.3	12.1
Dihydro-cer(d18:0/ 20:0)	6.4	0.6	5.8	-1.5	0.9	-2.0	-1.4	-6.1	-7.4
Dihydro-cer(d18:0/ 22:0)	3	0.7	9.9	-7.5	-2	-2.9	0.2	-6.1	-2.6
Dihydro-cer(d18:0/ 24:0)	-0.03	0.7	10.3	-6.5	-1.5	-2.9	-0.6	-5.5	-1.2
Dihydro-cer(d18:0/ 24:1)	5.3	0.6	7.7	-5.3	-0.9	-2.7	0.5	-6.3	-8.2
Cer(d18:1/ 16:0)	0.8	0.4	7.7	-2.8	0.1	28.3	-2.8	-5.1	-0.7
Cer(d18:1/ 18:0)	1.9	0.6	5	-3.1	-0.4	-1.0	-3.0	-4.3	-8.5
Cer(d18:1/ 20:0)	0.2	0.6	9	-1.7	0.8	-2.0	-1.8	-5.2	-2.6
Cer(d18:1/ 22:0)	-5.1	0.8	9.4	-4.5	-0.2	-2.3	-1.9	-5.1	0.2
Cer(d18:1/ 24:0)	-6.2	0.5	8.9	-4.5	-0.6	-3.5	-1.3	-4.6	-1.4
Cer(d18:1/ 24:1)	-3.8	0.6	6.9	-5	-1.2	-2.5	-1.6	-5.4	-5.9
Glucosyl-cer (d18:1/ 16:0)	-3.2	-0.3	-13.4	0.2	1.4	3.3	-2.0	1.7	-30.0
Glucosyl-cer (d18:1/ 18:0)	11.1	-1.1	-0.4	0.1	0.8	4.5	-4.3	-0.7	-16.6
Glucosyl-cer (d18:1/ 20:0)	2.4	-0.6	-0.1	-0.6	0.9	-0.4	-1.9	-3.5	-16.1
Glucosyl-cer (d18:1/ 22:0)	1.4	-0.4	8.4	-0.4	1.1	-0.5	-0.7	-2.6	-5.9
Glucosyl-cer (d18:1/ 24:0)	-3.1	-0.2	7.1	-0.2	1	0.2	-1.0	-3.3	-7.7
Glucosyl-cer (d18:1/ 24:1)	3.5	-0.2	5.3	-0.2	0.7	-1.3	-0.9	-3.1	-9.9
Dihydro-SM(d18:0, 16:0)	5.7	0	13.6	0	0.1	-3.7	1.5	-5.5	-16.5
Dihydro-SM(d18:0, 18:0)	7	-1.3	9.2	-1.3	0.8	-3.5	-0.5	-4.5	-2.9
Dihydro-SM(d18:0, 20:0)	4.3	-0.9	9.4	-0.9	1.1	-4.1	0.2	-4.3	-1.3
Dihydro-SM(d18:0, 22:0)	2.7	-1.9	12	-1.9	0.5	-4.5	-0.7	-5.1	1.1
Dihydro-SM(d18:0, 24:0)	1.1	-1.9	11.3	-1.9	0.6	-4.6	-0.2	-5.2	0.5
Dihydro-SM(d18:0, 24:1)	4.3	-1.1	9.1	-1.1	0.8	-5.1	0.3	-5.1	-2.7
SM(d18:1/ 16:0)	3	0	14.1	0	-0.3	-2.1	3.3	-4.8	-1.2
SM(d18:1/ 18:0)	6.6	0	13.6	0	0.2	-2.5	1.4	-3.9	0.2
SM(d18:1/ 20:0)	4.4	-0.7	9.9	-0.7	1	-4.0	0.3	-17.3	-1.7
SM(d18:1/ 22:0)	2	-1.2	11	-1.2	0.9	-4.6	-0.5	-4.4	1.1
SM(d18:1/ 24:0)	1.2	-0.9	12.6	-0.9	0.9	-6.9	-2.6	-7.5	0.7
SM(d18:1/ 24:1)	4.7	-0.7	9.8	-0.7	0.9	-4.7	0.8	-5.0	-2.2
Sphinganine	0.8	0	0.7	0	0.7	1.5	0.3	3.2	-21.7
Sphingosine	-3.2	-0.8	11.5	-0.8	0.8	-1.5	1.7	7.4	-10.8

(B) Odds ratio and 95%CI for variables changed more than 10% in the above analysis

	Hypertension	LCI	UCI	DM	LCI	UCI	Smoking	LCI	UCI	LDL-C	LCI	UCI	+All	LCI	UCI
Dihydro-cer(d18:0/ 16:0)	2.7	2.2	3.5	2.8	2.2	3.5	2.8	2.2	3.5	2.4	1.9	3.0	2.4	1.8	3.1
Dihydro-cer(d18:0/ 18:0)	2.7	2.2	3.5	2.8	2.2	3.5	2.7	2.1	3.4	2.7	2.1	3.4	2.0	1.6	2.5
Dihydro-cer(d18:0/ 20:0)	2.0	1.6	2.4	2.0	1.6	2.4	1.9	1.5	2.3	1.9	1.5	2.3	1.7	1.4	2.1
Dihydro-cer(d18:0/ 22:0)	2.1	1.7	2.6	2.1	1.7	2.7	2.0	1.6	2.5	1.9	1.6	2.4	1.7	1.4	2.2
Dihydro-cer(d18:0/ 24:0)	2.2	1.8	2.7	2.2	1.8	2.8	2.1	1.7	2.6	2.0	1.6	2.5	1.8	1.4	2.3
Dihydro-cer(d18:0/ 24:1)	2.3	1.9	2.9	2.4	1.9	3.0	2.2	1.8	2.8	2.2	1.8	2.8	2.0	1.6	2.6
Cer(d18:1/ 16:0)	3.0	2.0	4.5	2.2	1.8	2.8	2.2	1.8	2.7	2.1	1.7	2.7	1.8	1.5	2.3
Cer(d18:1/ 18:0)	2.3	1.8	2.8	2.2	1.8	2.8	2.2	1.8	2.7	2.2	1.8	2.7	2.0	1.6	2.5
Cer(d18:1/ 20:0)	2.0	1.6	2.5	2.0	1.6	2.5	1.9	1.6	2.4	1.9	1.5	2.3	1.7	1.3	2.1
Cer(d18:1/ 22:0)	1.8	1.5	2.2	1.8	1.5	2.2	1.7	1.4	2.1	1.6	1.3	2.0	1.4	1.2	1.8
Cer(d18:1/ 24:0)	2.0	1.7	2.5	2.1	1.7	2.6	2.0	1.6	2.5	1.9	1.6	2.4	1.7	1.4	2.2
Cer(d18:1/ 24:1)	2.2	1.8	2.8	2.3	1.8	2.8	2.2	1.8	2.7	2.1	1.7	2.7	1.9	1.5	2.4
Glucosyl-cer (d18:1/ 16:0)	1.0	0.8	1.2	1.0	0.8	1.1	1.0	0.8	1.2	1.1	0.9	1.3	1.1	0.9	1.4
Glucosyl-cer (d18:1/ 18:0)	1.4	1.2	1.7	1.3	1.1	1.5	1.3	1.1	1.6	1.3	1.1	1.6	1.3	1.0	1.6
Glucosyl-cer (d18:1/ 20:0)	2.0	1.6	2.4	1.9	1.6	2.4	1.9	1.6	2.4	2.0	1.6	2.5	1.9	1.5	2.4
Glucosyl-cer (d18:1/ 22:0)	1.8	1.5	2.3	1.8	1.5	2.3	1.8	1.5	2.2	1.7	1.4	2.1	1.6	1.2	2.0
Glucosyl-cer (d18:1/ 24:0)	1.7	1.4	2.1	1.7	1.4	2.1	1.7	1.4	2.1	1.6	1.3	2.0	1.5	1.2	1.9
Glucosyl-cer (d18:1/ 24:1)	2.2	1.8	2.7	2.2	1.8	2.7	2.1	1.7	2.7	2.1	1.7	2.6	1.9	1.5	2.4
Dihydro-SM(d18:0, 16:0)	2.0	1.6	2.4	2.1	1.7	2.6	1.9	1.6	2.4	1.8	1.4	2.2	1.9	1.5	2.5
Dihydro-SM(d18:0, 18:0)	2.2	1.8	2.8	2.3	1.9	2.9	2.2	1.8	2.8	2.1	1.7	2.6	1.9	1.5	2.4
Dihydro-SM(d18:0, 20:0)	1.8	1.5	2.3	1.9	1.6	2.4	1.8	1.5	2.2	1.7	1.4	2.1	1.5	1.2	1.9
Dihydro-SM(d18:0, 22:0)	2.5	2.0	3.1	2.6	2.0	3.2	2.4	2.0	3.1	2.3	1.8	2.9	2.0	1.6	2.5
Dihydro-SM(d18:0, 24:0)	2.3	1.9	2.9	2.4	2.0	3.1	2.3	1.9	2.9	2.2	1.7	2.8	1.9	1.5	2.4
Dihydro-SM(d18:0, 24:1)	2.5	2.0	3.2	2.7	2.1	3.4	2.5	2.0	3.2	2.4	1.9	3.1	2.1	1.7	2.7
SM(d18:1/ 16:0)	1.9	1.6	2.4	2.0	1.7	2.5	1.9	1.5	2.3	1.7	1.4	2.1	1.6	1.3	2.0
SM(d18:1/ 18:0)	1.6	1.3	2.0	1.7	1.4	2.0	1.6	1.3	1.9	1.4	1.2	1.8	1.3	1.1	1.6
SM(d18:1/ 20:0)	1.9	1.6	2.4	2.0	1.7	2.5	1.7	1.4	2.0	1.8	1.5	2.3	1.6	1.3	2.1
SM(d18:1/ 22:0)	2.1	1.7	2.6	2.2	1.8	2.7	2.1	1.7	2.6	2.0	1.6	2.4	1.7	1.4	2.2
SM(d18:1/ 24:0)	2.3	1.9	2.9	2.4	1.0	1.1	2.3	1.9	2.9	2.2	1.8	2.8	1.9	1.5	2.5
SM(d18:1/ 24:1)	2.3	1.8	2.9	2.4	1.9	3.0	2.3	1.8	2.9	2.2	1.7	2.7	1.9	1.5	2.5
Sphinganine	1.6	1.3	2.0	1.6	1.3	2.0	1.6	1.3	1.9	1.6	1.3	2.0	1.6	1.3	2.1
Sphingosine	3.4	2.6	4.7	3.4	2.6	4.6	3.2	2.4	4.4	3.1	2.3	4.2	2.8	2.1	3.9

Supplemental Table 3. Effect modification of results from the logistic regression analyses of lipid markers on coronary artery disease (CAD). (A) We generated an interaction term between the lipid marker (left) and the variable (top) to generate p-heterogeneity values. If the interaction term was significant in the model, we conducted analyses stratified by the effect modifier. (B) normotensive and (C) hypertensive, showing unadjusted OR (95%CI) (UA OR) and minimally-adjusted OR (95% CI) (age, sex, BMI) (A OR) with lower (LCI) and upper (UCI) CI displayed as well. DM, diabetes mellitus, MI, myocardial infarction, CERT1, ceramide risk score, SIC, sphingolipid inclusive CAD risk score, LDL-C, low density lipoprotein cholesterol.

(A) Table of lipid by variable interaction terms showing p-heterogeneity with significance denoted by an asterisk

	Age	Sex	BMI	DM	hypertension
Cer (d18:1/16:0)	0.557	0.672	0.098	0.002	1.82E-05*
Cer (d18:1/18:0)	0.282	0.347	0.140	0.001	3.57E-06*
Cer (d18:1/24:1)	0.285	0.675	0.130	0.000	8.82E-06*
CERT1	0.760	0.810	0.913	0.004	9.35E-07*
SIC	0.161	0.490	0.277	0.004	1.02E-05*
LDL-C	0.820	0.725	0.119	6.84E-12*	7.71E-10*

(B) OR and 95% CI for CAD for normotensive participants

Factor	UA OR	LCI	UCI	A OR	LCI	UCI
Dihydro-cer(d18:0/ 16:0)	2.33	1.82	3.03	2.43	1.86	3.24
Dihydro-cer(d18:0/ 18:0)	2.62	2.00	3.51	2.80	2.09	3.85
Dihydro-cer(d18:0/ 20:0)	1.79	1.41	2.29	1.97	1.53	2.58
Dihydro-cer(d18:0/ 22:0)	2.12	1.66	2.75	2.31	1.77	3.08
Dihydro-cer(d18:0/ 24:0)	2.26	1.76	2.95	2.34	1.79	3.11
Dihydro-cer(d18:0/ 24:1)	2.37	1.84	3.13	2.65	2.00	3.59
Cer(d18:1/ 16:0)	2.15	1.69	2.79	2.15	1.66	2.84
Cer(d18:1/ 18:0)	2.25	1.76	2.93	2.34	1.79	3.10
Cer(d18:1/ 20:0)	2.04	1.60	2.64	2.05	1.58	2.69
Cer(d18:1/ 22:0)	1.79	1.42	2.28	1.70	1.34	2.20
Cer(d18:1/ 24:0)	2.12	1.66	2.73	2.00	1.55	2.61
Cer(d18:1/ 24:1)	2.38	1.85	3.12	2.30	1.76	3.06
Glucosyl-cer (d18:1/ 16:0)	1.07	0.87	1.33	1.04	0.82	1.32
Glucosyl-cer (d18:1/ 18:0)	1.20	0.97	1.49	1.34	1.06	1.71
Glucosyl-cer (d18:1/ 20:0)	2.03	1.60	2.61	2.07	1.60	2.72
Glucosyl-cer (d18:1/ 22:0)	1.91	1.50	2.46	1.91	1.47	2.51
Glucosyl-cer (d18:1/ 24:0)	1.86	1.47	2.38	1.81	1.41	2.36
Glucosyl-cer (d18:1/ 24:1)	2.11	1.65	2.75	2.20	1.69	2.91
Dihydro-SM(d18:0, 16:0)	1.76	1.40	2.24	1.89	1.47	2.46
Dihydro-SM(d18:0, 18:0)	2.13	1.67	2.77	2.34	1.80	3.12
Dihydro-SM(d18:0, 20:0)	1.68	1.34	2.14	1.78	1.39	2.31
Dihydro-SM(d18:0, 22:0)	2.35	1.82	3.10	2.46	1.88	3.30
Dihydro-SM(d18:0, 24:0)	2.28	1.77	2.99	2.36	1.80	3.15
Dihydro-SM(d18:0, 24:1)	2.39	1.84	3.16	2.55	1.93	3.46
SM(d18:1/ 16:0)	1.76	1.40	2.26	1.82	1.41	2.38
SM(d18:1/ 18:0)	1.43	1.15	1.80	1.56	1.23	2.00
SM(d18:1/ 20:0)	1.76	1.40	2.24	1.87	1.46	2.43
SM(d18:1/ 22:0)	1.95	1.54	2.50	2.01	1.56	2.62
SM(d18:1/ 24:0)	2.23	1.73	2.92	2.25	1.72	3.01
SM(d18:1/ 24:1)	2.11	1.65	2.74	2.25	1.72	2.99
Sphinganine	1.68	1.33	2.15	1.64	1.27	2.15
Sphingosine	3.51	2.55	4.97	3.42	2.43	4.97

(C) OR and 95% CI for CAD for hypertensive participants

Factor	UA OR	LCI	UCI	A OR	LCI	UCI
Dihydro-cer(d18:0/ 16:0)	2.58	1.84	3.73	2.53	1.79	3.70
Dihydro-cer(d18:0/ 18:0)	2.34	1.66	3.41	2.48	1.73	3.66
Dihydro-cer(d18:0/ 20:0)	1.91	1.38	2.71	1.98	1.40	2.89
Dihydro-cer(d18:0/ 22:0)	1.65	1.21	2.30	1.71	1.24	2.40
Dihydro-cer(d18:0/ 24:0)	1.86	1.35	2.63	1.88	1.35	2.69
Dihydro-cer(d18:0/ 24:1)	1.81	1.32	2.54	1.91	1.36	2.73
Cer(d18:1/ 16:0)	2.25	1.63	3.19	2.33	1.65	3.37
Cer(d18:1/ 18:0)	2.05	1.50	2.85	2.16	1.55	3.07
Cer(d18:1/ 20:0)	1.84	1.35	2.55	1.91	1.37	2.70
Cer(d18:1/ 22:0)	1.86	1.36	2.60	1.86	1.34	2.65
Cer(d18:1/ 24:0)	2.19	1.59	3.10	2.12	1.52	3.04
Cer(d18:1/ 24:1)	2.09	1.52	2.92	2.10	1.51	2.99
Glucosyl-cer (d18:1/ 16:0)	1.01	0.76	1.36	0.95	0.71	1.30
Glucosyl-cer (d18:1/ 18:0)	1.51	1.11	2.06	1.62	1.18	2.26
Glucosyl-cer (d18:1/ 20:0)	1.87	1.36	2.64	1.83	1.32	2.60
Glucosyl-cer (d18:1/ 22:0)	1.79	1.31	2.52	1.76	1.26	2.52
Glucosyl-cer (d18:1/ 24:0)	1.78	1.30	2.49	1.65	1.19	2.33
Glucosyl-cer (d18:1/ 24:1)	2.27	1.62	3.26	2.19	1.54	3.20
Dihydro-SM(d18:0, 16:0)	2.13	1.54	3.02	2.12	1.52	3.04
Dihydro-SM(d18:0, 18:0)	1.90	1.39	2.65	2.02	1.45	2.88
Dihydro-SM(d18:0, 20:0)	1.85	1.36	2.55	1.89	1.37	2.65
Dihydro-SM(d18:0, 22:0)	2.37	1.67	3.47	2.41	1.68	3.57
Dihydro-SM(d18:0, 24:0)	2.29	1.62	3.34	2.27	1.59	3.34
Dihydro-SM(d18:0, 24:1)	2.42	1.72	3.49	2.45	1.72	3.60
SM(d18:1/ 16:0)	2.12	1.53	3.01	2.10	1.49	3.03
SM(d18:1/ 18:0)	1.63	1.21	2.22	1.69	1.24	2.33
SM(d18:1/ 20:0)	1.98	1.45	2.75	2.03	1.47	2.88
SM(d18:1/ 22:0)	2.21	1.60	3.13	2.22	1.59	3.19
SM(d18:1/ 24:0)	2.63	1.84	3.89	2.63	1.81	3.96
SM(d18:1/ 24:1)	2.30	1.65	3.27	2.32	1.65	3.37
Sphinganine	1.56	1.12	2.22	1.73	1.21	2.55
Sphingosine	3.12	1.98	5.20	3.30	2.05	5.61

Supplemental Table 4. To compare machine learning variable reduction techniques to classical variable reduction techniques, we performed a stepwise (forwards and backwards) logistic regression. The input included age, sex, BMI, diabetes, hypertension, smoking, LDL-C, VLDL-C, HDL-C, total-C, triglycerides, and the 32 sphingolipid species. The table below depicts the variables retained in the model in addition to their standard errors and p-values.

	Estimate	Standard Error	P-Value
Intercept	0.670	0.530	0.210
Age	0.007	0.002	2.48E-04
Sex	0.188	0.031	1.40E-09
BMI	-0.004	0.003	0.101
Diabetes	0.178	0.037	2.53E-06
Hypertension	0.133	0.028	1.99E-06
Smoking	0.046	0.028	0.104
LDL-C	0.003	0.000	3.76E-15
Sphinganine	-0.177	0.086	0.040
SM(d18:1/18:0)	-1.036	0.168	1.33E-09
SM(d18:0/20:0)	-0.825	0.219	1.83E-04
SM(d18:1/20:0)	0.557	0.234	0.018
SM(d18:1/22:0)	-0.490	0.205	0.017
SM(d18:0/22:0)	0.616	0.230	0.008
GlcCer (d18:1/24:1)	0.148	0.072	0.041
SM(d18:0/24:0)	-0.511	0.263	0.052
SM(d18:0/24:0)	0.700	0.182	1.33E-04
Cer(d18:1/22:0)	-1.077	0.203	1.47E-07
Cer(d18:0/24:0)	-0.453	0.150	0.003
Cer(d18:1/24:0)	1.343	0.228	5.99E-09
Cer(d18:1/18:0)	0.479	0.106	7.34E-06
Cer(d18:0/16:0)	0.270	0.115	0.019
Cer(d18:0/18:0)	0.200	0.098	0.043
Sphingosine	0.203	0.064	0.002

Supplemental Table 5. Least Absolute Shrinkage and Selection Operator (LASSO) regression performed on sphingolipids and conventional coronary artery disease (CAD) lipid markers. (A) LASSO selected variables (Lipid variable) with their average coefficient (Coefficient) and frequency of selection (Frequency) over the 10 iterations. The average accuracy of case and control detection across all 10 iterations was 84.6%. Accuracy is defined as the percentage of cases and controls correctly identified.

(A)

Lipid variable	Coefficient	Frequency
logCer(d18:1/24:0)	4.686309497	10
logCer(d18:1/18:0)	3.024266512	10
logSM(d18:0/24:1)	2.733016943	10
logSM(d18:1/24:0)	2.437666391	10
logCer(d18:0/18:0)	1.993895534	10
logCer(d18:1/24:1)	1.810724709	10
logSM(d18:0/22:0)	1.537397593	5
logGlcCer (d18:1/24:1)	1.525020383	10
logSM(d18:1/16:0)	1.193604832	7
logSphingosine	1.162494717	10
logSM(d18:1/20:0)	0.862010913	1
logCer(d18:0/16:0)	0.782965718	10
logCer(d18:1/16:0)	0.48397109	4
logGlcCer (d18:1/20:0)	0.278394995	6
logCer(d18:1/20:0)	0.021446838	2
logSM(d18:0/24:0)	-0.001197061	1
logSM(d18:1/24:1)	-0.415124981	1
logGlcCer (d18:1/16:0)	-0.477480141	5
logCer(d18:0/22:0)	-0.515820522	9
logGlcCer (d18:1/24:0)	-0.600200334	3
logGlcCer (d18:1/18:0)	-0.630045642	7
logSM(d18:0/16:0)	-0.634674099	1
logCer(d18:0/20:0)	-0.657002555	10
logSphinganine	-0.70663971	10
logSM(d18:0/18:0)	-0.724004213	1
logGlcCer (d18:1/22:0)	-0.764175824	4
logCer(d18:0/24:1)	-1.112942293	2
logCer(d18:0/24:0)	-1.583361189	7
logSM(d18:1/22:0)	-2.723032113	2
logSM(d18:0/20:0)	-4.638343787	10
logCer(d18:1/22:0)	-4.681281021	10
logSM(d18:1/18:0)	-6.622572894	10

Supplemental Table 6. Random Forest (RF) performed on sphingolipids and sphingolipids with conventional coronary artery disease (CAD) lipid markers. For the input including only sphingolipids, (A) the top ten RF selected variables (Lipid variable) for each iteration with their frequency of selection (Frequency) over the 5 iterations are displayed. The average accuracy of case and control detection across the 5 RF runs for the sphingolipid only input was 81.5%. (B) is the shows the same data as (A) but for the sphingolipid and standard clinical lipid (Total-C, LDL-C, VLDL-C, HDL-C, triglyceride) input. The average accuracy of case and control detection across the 5 RF runs for the sphingolipids and standard clinical lipid input was 82.33%. For RF analysis the default was 500 decision trees.

(A)

Lipid Variable	Frequency
Cer(d18:0/16:0)	5
Sphingosine	5
Cer(d18:0/18:0)	4
SM(d18:1/18:0)	5
SM(d18:0/24:1)	5
SM(d18:1/24:0)	3
GlcCer(d18:1/24:0)	1
Cer(d18:1/18:0)	2
SM(d18:0/22:0)	4
Cer(d18:0/24:0)	3
Cer(d18:1/16:0)	4
SM(d18:1/24:1)	2
Cer(d18:1/24:1)	2
GlcCer(d18:1/20:0)	2
Cer(d18:0/20:0)	1
SM(d18:1/16:0)	1
SM(d18:1/22:0)	1

(B)

Lipid Variable	Frequency
LDL-C	5
Cer(d18:0/16:0)	5
Cer(d18:0/18:0)	5
Sphingosine	5
Cer(d18:0/20:0)	3
GlcCer(d18:1/20:0)	3
SM(d18:1/18:0)	5
Cer(d18:0/24:1)	1
Cer(d18:1/16:0)	3
Cer(d18:1/24:1)	4
SM(d18:1/24:0)	1
SM(d18:0/24:1)	3
total-C	2
VLDL-C	1
Cer(d18:1/18:0)	2
GlcCer(d18:1/24:1)	1
SM(d18:1/22:0)	1

Supplemental Table 7. Net reclassification index (NRI) and integrative discriminatory index (IDI) to compare the American Heart Association (AHA) and American College of Cardiology (ACC) based guidelines (age, sex, BMI, diabetes, hypertension, smoking, LDL-C, HDL-C, VLDL-C, total-C, triglycerides) to an AHA/ACC guideline model with the addition of (A) the cardiac event risk test (CERT1) and (B) the sphingolipid inclusive CAD (SIC) score. Continuous NRI and IDI are displayed with 95% confidence intervals (CI). NRI values above 0.6 should be conisdered strong, those above 0.4 intermediate, and those below 0.2 weak. A reclassification table showing upward and downward movement with the new model.

(A) AHA/ACC model versus AHA/ACC + CERT1

	CAD Cases	Controls
Down	142	192
Unchanged	0	0
Up	70	254

Continuous NRI	0.48	95% CI (0.32-0.64)
IDI	0.04	95% CI (0.03-0.06)

(B) AHA/ACC model versus AHA/ACC + SIC

	CAD Cases	Controls
Down	160	188
Unchanged	0	0
Up	52	258

Continuous NRI	0.67	95% CI (0.52-0.81)
IDI	0.1	95% CI (0.08-0.11)

Supplemental Table 8. Results from the primary logistic regression of coronary artery disease (CAD) on sphingolipid species. (A) minimally-adjusted model (age, sex, BMI); (B) multivariable-adjusted model including age, sex, BMI, total-C, LDL-C, VLDL-C, triglycerides, hypertension, diabetes, smoking. Associations considered significant at a false discovery rate (FDR) of 0.05. BH, Benjamini Hochberg.

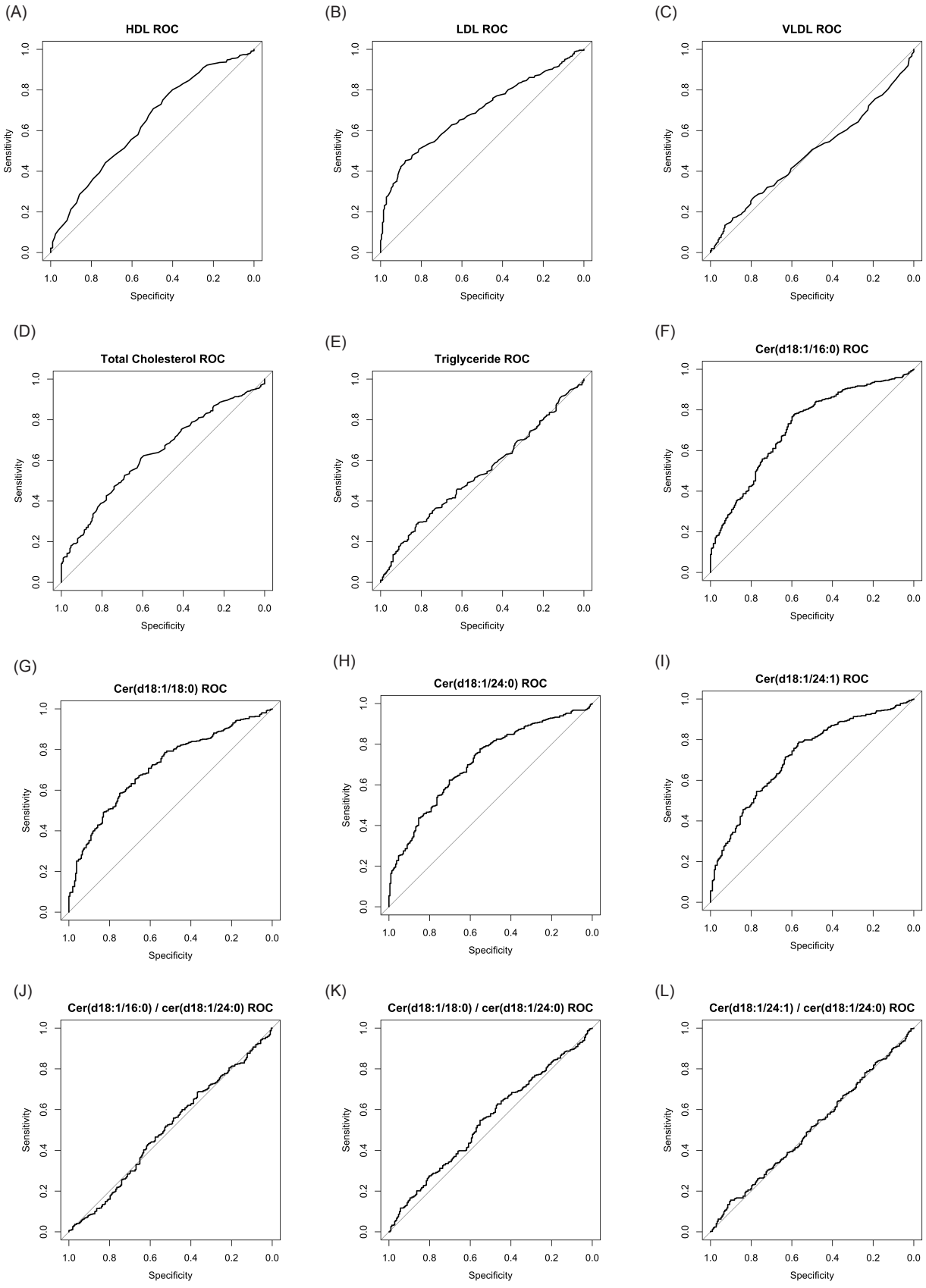
(A) Minimally-adjusted

lipid	p-value	BH	FDR<0.05
Dihydro Cer d18:0, 16:0	2.00E-16	1.60E-15	significant
Dihydro Cer d18:0, 18:0	2.00E-16	1.60E-15	significant
Dihydro Cer d18:0, 20:0	2.36E-10	5.03E-10	significant
Dihydro Cer d18:0, 22:0	1.39E-12	4.04E-12	significant
Dihydro Cer d18:0, 24:0	1.29E-14	4.59E-14	significant
Dihydro Cer d18:0, 24:1	4.26E-15	1.70E-14	significant
Cer d18:1, 16:0	5.83E-16	3.11E-15	significant
Cer d18:1, 18:0	5.40E-16	3.11E-15	significant
Cer d18:1, 20:0	6.12E-13	1.96E-12	significant
Cer d18:1, 22:0	3.68E-11	9.81E-11	significant
Cer d18:1, 24:0	1.61E-15	7.36E-15	significant
Cer d18:1, 24:1	2.00E-16	1.60E-15	significant
Gcs Cer d18:1, 16:0	9.75E-01	9.75E-01	not significant
Gcs Cer d18:1, 18:0	3.04E-01	3.14E-01	not significant
Gcs Cer d18:1, 20:0	3.34E-08	4.28E-08	significant
Gcs Cer d18:1, 22:0	2.87E-08	3.83E-08	significant
Gcs Cer d18:1, 24:0	8.14E-08	1.00E-07	significant
Gcs Cer d18:1, 24:1	5.11E-10	9.62E-10	significant
SM d18:0, 16:0	1.22E-09	2.05E-09	significant
SM d18:0, 18:0	2.72E-09	3.96E-09	significant
SM d18:0, 20:0	1.25E-07	1.43E-07	significant
SM d18:0, 22:0	9.08E-10	1.61E-09	significant
SM d18:0, 24:0	1.57E-10	3.59E-10	significant
SM d18:0, 24:1	1.40E-10	3.45E-10	significant
SM d18:1, 16:0	3.77E-10	7.54E-10	significant
SM d18:1, 18:0	2.54E-06	2.80E-06	significant
SM d18:1, 20:0	1.03E-07	1.22E-07	significant
SM d18:1, 22:0	4.28E-09	5.95E-09	significant
SM d18:1, 24:0	1.44E-09	2.30E-09	significant
SM d18:1, 24:1	1.85E-09	2.82E-09	significant
Sphinganine	4.61E-06	4.92E-06	significant
Sphingosine	2.00E-16	1.60E-15	significant

(B) Multivariate-adjusted model

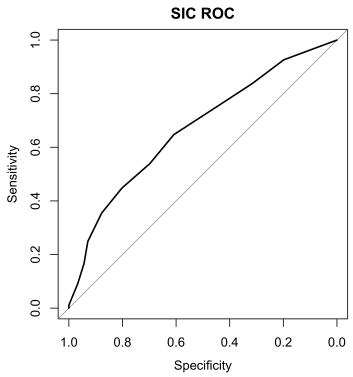
lipid	p-value	BH	FDR<0.05
Dihydro Cer d18:0, 16:0	2.10E-09	3.36E-08	significant
Dihydro Cer d18:0, 18:0	5.12E-10	1.64E-08	significant
Dihydro Cer d18:0, 20:0	1.67E-05	2.67E-05	significant
Dihydro Cer d18:0, 22:0	9.04E-06	1.61E-05	significant
Dihydro Cer d18:0, 24:0	2.39E-06	4.74E-06	significant
Dihydro Cer d18:0, 24:1	3.54E-08	2.83E-07	significant
Cer d18:1, 16:0	2.52E-06	4.74E-06	significant
Cer d18:1, 18:0	9.56E-08	6.12E-07	significant
Cer d18:1, 20:0	6.13E-05	8.53E-05	significant
Cer d18:1, 22:0	4.68E-03	4.99E-03	significant
Cer d18:1, 24:0	2.08E-05	3.17E-05	significant
Cer d18:1, 24:1	4.86E-07	1.29E-06	significant
Gcs Cer d18:1, 16:0	4.34E-01	4.34E-01	not significant
Gcs Cer d18:1, 18:0	3.30E-02	3.40E-02	significant
Gcs Cer d18:1, 20:0	3.32E-07	9.66E-07	significant
Gcs Cer d18:1, 22:0	3.46E-04	4.26E-04	significant
Gcs Cer d18:1, 24:0	1.02E-03	1.13E-03	significant
Gcs Cer d18:1, 24:1	1.16E-07	6.19E-07	significant
SM d18:0, 16:0	9.04E-05	1.21E-04	significant
SM d18:0, 18:0	5.91E-07	1.35E-06	significant
SM d18:0, 20:0	3.81E-04	4.52E-04	significant
SM d18:0, 22:0	1.80E-07	7.20E-07	significant
SM d18:0, 24:0	7.87E-07	1.68E-06	significant
SM d18:0, 24:1	1.57E-08	1.67E-07	significant
SM d18:1, 16:0	1.01E-04	1.29E-04	significant
SM d18:1, 18:0	1.40E-07	6.40E-07	significant
SM d18:1, 20:0	6.03E-05	8.53E-05	significant
SM d18:1, 22:0	1.40E-05	2.36E-05	significant
SM d18:1, 24:0	5.26E-07	1.29E-06	significant
SM d18:1, 24:1	2.71E-07	8.67E-07	significant
sphinganine	5.58E-04	6.38E-04	significant
sphingosine	2.49E-07	8.67E-07	significant

Supplemental Figure 1. Receiver operating characteristic (ROC) curves for discrimination between coronary artery disease (CAD) cases and controls by lipids in the Utah CAD study. (A) high density lipoprotein (HDL) Cholesterol, (B) low density lipoprotein (LDL) Cholesterol, (C) very low density lipoprotein (VLDL) cholesterol, (D) total cholesterol, (E) triglycerides, (F) Cer(d18:1/16:0), (G) Cer(d18:1/18:0), (H) Cer(d18:1/24:0), (I) Cer(d18:1/24:1)

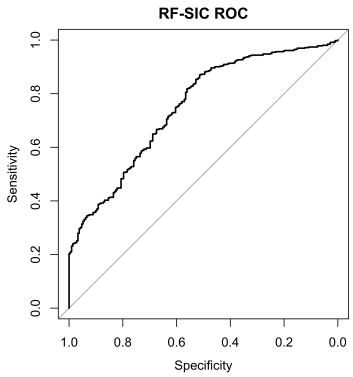


Supplemental Figure 2. ROC curves for (J) Cer(d18:1/16:0) / Cer(d18:1/24:0) ratio, (K) Cer(d18:1/18:0) / Cer(d18:1/24:0) ratio (L) Cer(d18:1/24:1) / Cer(d18:1/24:0) ratio, (M) CERT1 score, (N)RF-SIC score, (O) RF-SIC+, (P) LASSO-SIC score, (Q) LASSO-SIC2 score, (R) LASSO-SIC3 score, (S) SIC score. (T) shows all the C-statistics for each ROC curve.

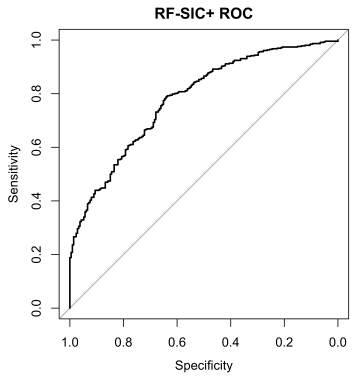
(M)



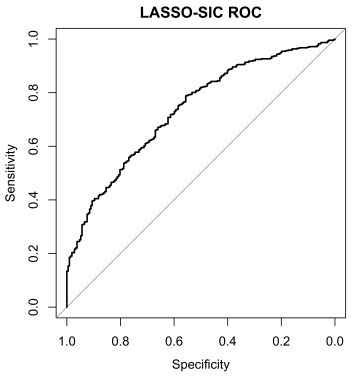
(N)



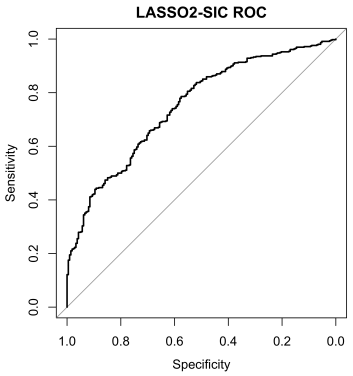
(O)



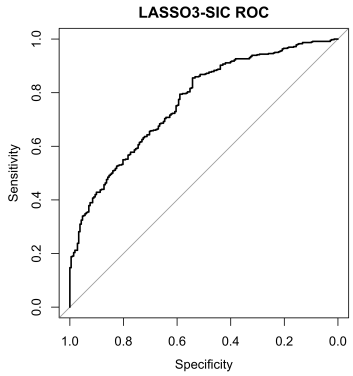
(P)



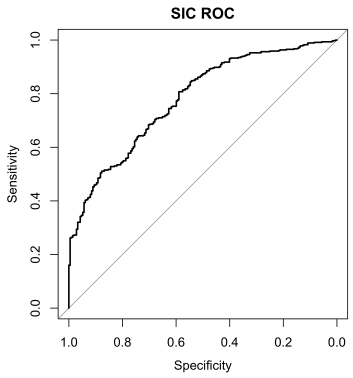
(Q)



(R)



(S)



(T)

Clinical Index	C-statistic
HDL-C	0.63
LDL-C	0.69
VLDL-C	0.49
Total-C	0.64
Triglycerides	0.54
Cer(d18:1/16:0)	0.71
Cer(d18:1/18:0)	0.71
Cer(d18:1/1240)	0.71
Cer(d18:1/24:1)	0.72
Cer(d18:1/16:0) / Cer(d18:1/24:0)	0.5
Cer(d18:1/18:0) / Cer(d18:1/24:0)	0.54
Cer(d18:1/24:1) / Cer(d18:1/24:0)	0.51
CERT1	0.67
RF-SIC	0.75
RF-SIC+	0.78
LASSO-SIC	0.73
LASSO-SIC2	0.75
LASSO-SIC3	0.76
SIC	0.79

Supporting Materials for 131838-JCI-CMED-DN-2

This is an observational study, as the assignment of the medical intervention was not at the discretion of the investigators. We opted not to conduct a new trial registration for this already closed study. We provide the following supportive information.

STROBE Statement—checklist of items

	Item No	Recommendation
Title and abstract	1	(*a*) Title: Machine Learning Reveals Serum Sphingolipids as Cholesterol-Independent Biomarkers of Coronary Artery Disease
		(*b*) 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.
Introduction		
Background/rationale	2	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. 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. 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.

Objectives	3	The goals of this study were to identify environmental and genetic determinants of early onset familial coronary artery disease. The study has been ongoing since 1977, but is now closed to additional recruitment. The mechanisms for recruitment and screening of participants changed little during its course. Probands and their relatives were invited to our screening clinic, completed medical history questionnaires, had blood drawn, and underwent clinical measurements. Blood and DNA samples were stored. Current efforts are directed at conducting lipidomic assessments of the samples to identify new correlates of disease.
Methods		
Study design	4	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).
Setting	5	Cases were recruited between 1977 and 2000 from Intermountain Healthcare discharge records or the Family Health Tree Program in Utah (58). Both case and control populations were selected from the same source population of Salt Lake City, Utah.
Participants	6	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. Cases had similar age of onset to at least one first degree relative (parent, sibling, or child. Controls representative of the Utah population were randomly sampled from 1980-1986 from (i) the parents of students participating in the Family Health Tree Program, a study of family health among Utah high schools; and (ii) spouse pairs participating in a study on psychological factors concerning CAD. 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.
Variables	7	As noted above, cases were defined by the original study recruitment criteria as myocardial infarction (MI), percutaneous transluminal coronary angioplasty (PCTA), or coronary artery bypass grafting (CABG). Measured endpoints included available clinical data and new sphingolipidomic assessments.
Data sources/ measurement	8	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/m2), 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 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. Lipoprotein concentrations were measured using a microscale ultracentrifugation method. Serum samples were aliquoted and stored at -80 °C. The collection laboratory participates in the Centers for Disease Control Lipid Standardization Program. Lipidomics were done by mass spectroscopy. 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.
Bias	9	Patient recruitment occurred several decades ago and was representative of the Utah population at that time. We applied unbiased machine learning to identify sphingolipids associated with CAD.
Study size	10	Sample size was based on available, banked samples
Quantitative variables	11	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. To calculate the Ceramide Risk Score (CERT1) that is in clinical use, 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 4th quartile, 1 point to the 3rd quartile, and 0 points to the bottom two quartiles, with total CERT1 scores ranging from 0-12.
Statistical methods	12	Participant characteristics were summarized as mean ± standard deviation for continuous variables or N (%) for categorical variables. 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 and were log10 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 modelling. 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. Where effect modification was present (p-value for the interaction term<0.05), 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.

We applied machine learning 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. We then performed Least Absolute Shrinkage and Selection Operator (LASSO) regression and Random Forest analysis. 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. 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. 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).

All analyses were performed in R 3.5.1. Associations were considered statistically significant at a false discovery rate (FDR) <0.05 to control for multiple statistical tests.

Results

Participants	13	n=462 cases and n=212 controls

Descriptive data	14	(a) We quantified 32 sphingolipids. All sphingolipids measured, excepting two glucosylceramides, were elevated in CAD cases compared with controls. We provide 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.
		(b) Indicate number of participants with missing data for each variable of interest: 0
Main results	16	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 ORperSD 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.
Discussion		
Key results	18	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.
Limitations	19	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). Third, this study lacks a validation cohort for the novel SIC score. 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.
Interpretation	20	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.

Generalisability	21	Follow-up studies will be needed to determine whether these findings are broadly generalizable
Other information		
Funding	22	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).