ORIGINAL RESEARCH

Spectrum of Apolipoprotein AI and Apolipoprotein AII Proteoforms and Their Associations With Indices of Cardiometabolic Health: The CARDIA Study

John T. Wilkins 🔟, MD, MS*; Henrique S. Seckler, MS*; Jonathan Rink 🔟, PhD; Philip D. Compton, PhD; Luca Fornelli 🔟, PhD; C. Shad Thaxton, MD, PhD; Rich LeDuc 🔟, PhD; David Jacobs 🔟, PhD; Peter F. Doubleday, PhD; Allan Sniderman 🔟, MD; Donald M. Lloyd-Jones 🔟, MD, MSc; Neil L. Kelleher, PhD

BACKGROUND: ApoAI (apolipoproteins AI) and apoAII (apolipoprotein AII) are structural and functional proteins of high-density lipoproteins (HDL) which undergo post-translational modifications at specific residues, creating distinct proteoforms. While specific post-translational modifications have been reported to alter apolipoprotein function, the full spectrum of apoAI and apoAII proteoforms and their associations with cardiometabolic phenotype remains unknown. Herein, we comprehensively characterize apoAI and apoAII proteoforms detectable in serum and their post-translational modifications and quantify their associations with cardiometabolic health indices.

METHODS AND RESULTS: Using top-down proteomics (mass-spectrometric analysis of intact proteins), we analyzed paired serum samples from 150 CARDIA (Coronary Artery Risk Development in Young Adults) study participants from year 20 and 25 exams. Measuring 15 apoAI and 9 apoAII proteoforms, 6 of which carried novel post-translational modifications, we quantified associations between percent proteoform abundance and key cardiometabolic indices. Canonical (unmodified) apoAI had inverse associations with HDL cholesterol and HDL-cholesterol efflux, and positive associations with obesity indices (body mass index, waist circumference), and triglycerides, whereas glycated apoAI showed positive associations with serum glucose and diabetes mellitus. Fatty-acid–modified ApoAI proteoforms had positive associations with HDL cholesterol and efflux, and inverse associations with obesity indices and triglycerides. Truncated and dimerized proteoforms of apoAII were associated with HDL cholesterol (positively) and obesity indices (inversely). Several proteoforms had no significant associations with phenotype.

CONCLUSIONS: Associations between apoAI and AII and cardiometabolic indices are proteoform-specific. These results provide "proof-of-concept" that precise chemical characterization of human apolipoproteins will yield improved insights into the complex pathways through which proteins signify and mediate health and disease.

Key Words: acylation ■ apolipoprotein AI ■ apolipoprotein AII ■ post-translational modifications ■ proteoform

ApoA1 (apolipoprotein AI) is the 11th most abundant protein in human serum.[1] ApoAI serves as a major structural scaffold and binding ligand for high-density lipoprotein (HDL) particles.[2,3] Thus, apoAI has a central role in lipid metabolism, mediating cholesterol transport and inflammatory, immunologic, and vasodilatory pathways.[4] Like HDL-cholesterol concentration (HDL-C), apoAI has strong associations with

<div style="border: 2px solid red; padding: 10px;">

CLINICAL PERSPECTIVE

What Is New?

- A proteoform is the precise molecular form of a protein, including allelic and splice variants, coding polymorphisms, and the nature, location(s) of any post-translational modifications.
- In this study, apolipoprotein AI (apoAI) and apolipoprotein AII (apoAII) present in human serum had 15 and 9 different proteoforms, respectively.
- The magnitude and direction of association of proteoform abundance to cardiometabolic characteristics of 150 individuals varied substantially across proteoforms and between proteoforms and total protein abundance.

What Are the Clinical Implications?

- Proteoform composition may vary significantly across biological states despite a smaller or undetectable difference in total protein concentration.
- Thus, proteoform-level characterization and quantification of apolipoproteins may enhance clinical and biological inference obtained from the measurement of total protein concentration alone.

</div>

<div style="border: 2px solid red; padding: 10px;">

Nonstandard Abbreviations and Acronyms

ApoAI	apolipoprotein AI
ApoAII	apolipoprotein AII
ApoB	apolipoprotein B
CARDIA	Coronary Artery Risk Development in Young Adults study
K88acylApoAI	proteoforms characterized by acylation at lysine residue 88 (K88) of ApoAI
LC-MS	liquid chromatography/mass spectrometry
PTM	post-translational modification

</div>

markers of cardiometabolic health and risk of coronary heart disease (CHD).[5] Although it is unclear whether the relationship between apoAI and CHD risk is causal,[6,7] strong associations between apoAI concentration and metabolic characteristics that are causally related to CHD risk, such as apoB (apolipoprotein B) and insulin resistance, support the hypothesis that ApoAI may be a mechanistically important mediator of cardiometabolic health.[8,9]

The precise molecular form of a protein arising from a gene, including allelic and splice variants, coding polymorphisms, and the nature and location(s) of any post-translational modifications (PTMs), is called

a proteoform (Figure 1A).[10] Previous reports have described a limited number of apolipoprotein PTMs,[11] and several of those modifications are associated with differences in HDL functional properties.[12–16] However, these reports are sparse, and have traditionally relied on protein digestion and subsequent PTM detection on small peptides (ie, bottom-up proteomics), which forbids characterization of the precise proteoforms present in a sample.[10] Consequently, data on apolipoproteins proteoforms and on their associations with differences in phenotypes are limited

Recently, we have reported on a top-down proteomic methodology to quantify whole proteoforms of apoAI in human serum, designed to capture a more comprehensive picture of apolipoprotein molecular variation. From a pilot group of 8 samples obtained from CHAS (Chicago Healthy Aging Study), we observed that proteoforms characterized by acylation (covalent fatty-acid addition) at lysine residue 88 (K88) of ApoAI (K88acylApoAI) were positively associated with HDL efflux capacity.[12] An association between an apoAI proteoform and higher efflux capacity had not been described previously, as other reports of apoAI proteoforms and/or PTMs showed associations with lower HDL efflux.[15] In the CHAS study, however, we were unable to assess the associations between ApoAI proteoforms and other cardiometabolic characteristics because of the small sample size.

Thus, the objectives of the current study were to: characterize and quantify apoA-I proteoforms in serum samples obtained from 150 well characterized adults from the Coronary Artery Risk Development in Young Adults (CARDIA) cohort collected at 2 different exam cycles 5 years apart; and to quantify the associations between apoAI proteoforms and cardiometabolic health indices including HDL efflux capacity (Figure 1B). We hypothesized that specific apoAI proteoforms would exhibit differences in association with cardiometabolic characteristics, and, specifically, that K88acylApoAI proteoforms would be positively associated with HDL-C and HDL efflux and inversely associated with markers of poor cardiometabolic health.

Moreover, we set out to characterize the specificity of PTMs and overall proteoform chemistry and variation among apolipoproteins. We thus applied a similar proteoform characterization approach to apoAII (apolipoprotein A-II), the second most abundant HDL-associated protein. While some reports suggest apoAII is a mediator of HDL function[17,18] and that modifications of apoAII are associated differences in metabolism,[19] no comprehensive characterization of apoAII proteoforms and their associations with cardiometabolic characteristics had yet been performed. Therefore, the present study also reports on associations between CARDIA characteristics and proteoforms of apoAII.
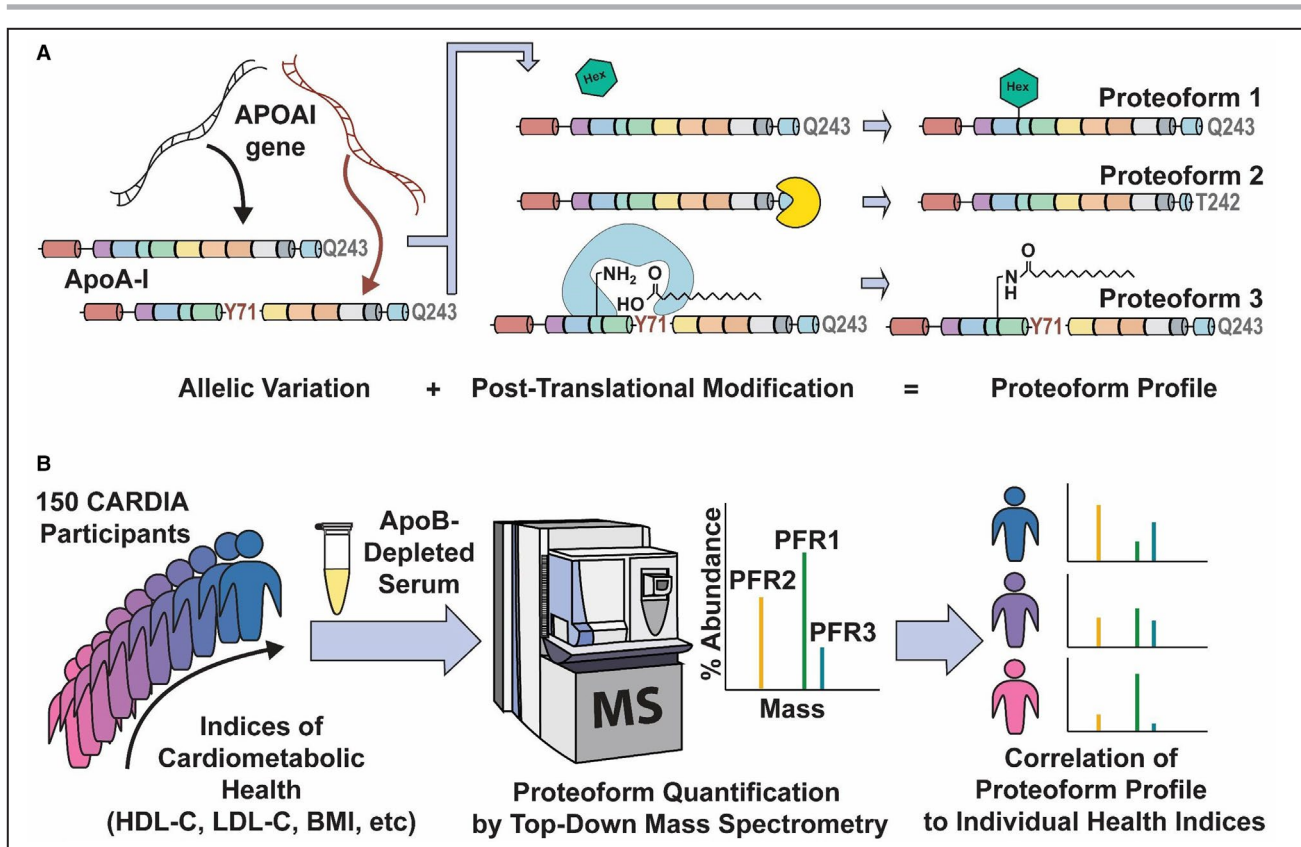
Figure 1. Apolipoprotein proteoform analysis.
A, The proteoform profile of a gene product. The translation of different backbones from a single gene, be it because of allelic variation (shown here with apoAI [apolipoprotein AI]) or splicing variation, combined with post-translational modifications to these backbones contribute to the creation of a panel of different chemical species, all products of the same gene. These species—and not necessarily the unmodified coding product of the gene—are the actual molecules circulating and carrying out biological function in an organism. Each of these chemical variants of a gene product is called a proteoform. **B**, Study design. To search for associations between proteoform profiles and cardiometabolic phenotype, we used serum samples from 150 previously phenotyped CARDIA participants. Two samples were analyzed for each participant, corresponding to years 20 and 25 of the CARDIA study. Targeted top-down mass spectrometry was used to discover, chemically characterize and quantify proteoforms of apoAI and apoAII. The percent contribution of each proteoform to the total amount of each apolipoprotein (ie, the quantitative proteoform profile) was compared with indices of cardiometabolic health for each individual studied. ApoAI indicates apolipoprotein AI; apoAII, apolipoprotein AII; apoB, apolipoprotein B; BMI indicates body mass index; CARDIA, Coronary Artery Risk Development in Young Adults; HDL-C, high-density lipoprotein cholesterol; and LDL-C, low-density lipoprotein cholesterol; PFR1−3, proteoform 1−3.

METHODS

Study Cohort

Samples were obtained from the CARDIA study. Details of the CARDIA cohort and methods for risk factor measurement have been described elsewhere.[20] Briefly, CARDIA is a community-based cohort, designed to study the development of cardiovascular disease risk factors and their clinical sequelae in Black and White young adults in the United States. Between 1985 and 1986, 5115 participants between ages 18 to 30 years, 51.5% Black, 54.5% women were enrolled. At baseline and at each subsequent examination, participants underwent extensive in-person measurement of CHD risk factors. Participants returned for examination at years 2, 5, 7, 10, 15, 20, 25, and 30, with 70% of the surviving cohort examined at year 30. In addition, health status and

hospitalizations for cardiovascular endpoints are collected annually outside of clinic visits. Risk factors were measured, and samples were stored for all follow-up examinations. At year 20, participants underwent coronary artery calcium (CAC) assessment. A total of 3547 participants returned for examination at year 20 and 3499 participants returned for examination at year 25. To date, the cohort has been followed through year 34.

Details of phenotypic characterization methodology, including CAC, HDL efflux, demographic and other serologic assessments are outlined in Data S1.

Sample Selection

CARDIA participants were eligible if they presented to the years 20 and 25 examinations, had serum samples available from both examinations, and had traditional risk factor measurement data, as well as CAC

at the year 20 examination. Our sampling approach (described in detail below) was designed to: (1) allow for detection of the maximal number of apoAI proteoforms, (2) allow for analysis of apoAI proteoform intensity across a range of HDL-C and efflux values, and (3) assess associations between proteoforms and subclinical atherosclerosis. We chose a sample size of 150 participants at years 20 and 25 (total number of samples: 300). Since this was discovery-mode analysis, we did not perform a power calculation.

Since throughput is limited using current top-down proteomics technology, we over-sampled outliers to maximize the probability that we would detect low frequency proteoforms and detect associations with phenotype. We chose a 2×3 sampling approach based on prevalent CAC at year 20 and HDL-C values at year 20 as outlined in Table 1. The cohort was stratified into each of the 6 categories, then we selected a random sample of 25 participants, who had sample available at years 20 and 25, from each CAC/HDL-C subgroup.

This study was approved by Northwestern Medicine's Institutional Review Board.

Incubation of ApoAI With Palmitic Acid

For detection of a potential non-enzymatic reaction between apoAI and free fatty acids, in-house-purified non-acylated apoAI was incubated with palmitic acid (Sigma-Aldrich, Saint Louis, MO) at serum-like conditions: (ApoAI)=1.5 g/L, (palmitate)=1.6 mmol/L, pH 7.4, 37°C. A 100 mmol/L ammonium bicarbonate buffer was used, for MS compatibility. Aliquots were collected at times: 2, 5, 24, and 48 hours, until apoAI oxidation/degradation prevented the potential detection of the acylated proteoform.

Proteoform Characterization and Quantification

Proteoforms were analyzed by liquid chromatography/mass spectrometry (LC-MS) of participant apoB-depleted serum samples. All peaks of similar charge, mass and retention time as apoAI or -all were characterized by a classic top-down proteomic approach. Serum samples of all individuals were run in randomized order in 6 LC-MS analysis blocks (2 collection years of 25 individuals per block). Proteoform intensities were

Table 1. The 2×3 Sampling Used to Select Participant Samples for Proteoform Analysis

	CAC=0	CAC >0
HDL-C >60 mg/dL	25	25
HDL-C 40–60 mg/dL	25	25
HDL-C <40 mg/dL	25	25

CAC indicates coronary artery calcium; and HDL-C, high-density lipoprotein cholesterol.

normalized by total ion current across runs and standardized across blocks. Standards of known apoAI concentration were analyzed along with samples in each block. Individuals' average percent abundance of each proteoform was compared with their characteristics in both CARDIA years 20 and 25. Associations between proteoforms and characteristics were assessed by either linear regression (for continuous characteristics) or t tests (for binary data). Then, either Pearson r and beta coefficients or fold differences were calculated, along with a p-score. A Benjamini-Hochberg multiple-test correction was done, and significance was asserted at a 5% false discovery rate (2-tailed type 1 error rate). Detailed descriptions of each of these analyses are present in the supplemental material.

RESULTS

Participant Characteristics

Characteristics of participants grouped by HDL-C levels are presented in Table 2. By sampling design, there were 50 in each HDL-C group, half of whom had CAC >0 AU. The mean HDL-C was 79, 49, and 34 mg/dL in the high, medium, and low HDL-C groups with CAC; and 73, 50, and 35 mg/dL in the same groups without CAC, respectively. When compared with the middle and low HDL-C groups, individuals in both high HDL-C groups were more likely to be women, and individuals with high HDL-C and CAC >0 were more likely to smoke. The high HDL-C group had lower BMI, waist circumference, LDL-C, triglycerides, and glucose levels. Individuals in higher HDL-C groups had higher HDL efflux capacity. There was no difference in ABCA-1-dependent efflux across groups.

Proteoform Spectrum of Serum ApoAI

We set out to quantitate the full spectrum of proteoforms of apoAI detectable by top-down LC-MS in participant sera. Figure S1 shows raw mass spectrometric data for species that were identified as apoA-I. LC-MS data revealed a total of 15 apoAI species of distinguishable mass which were commonly detected across the individuals analyzed. Supporting fragmentation data can be found on the MassIVE database for this project, along with confidence metrics for proteoform identification (massive.ucsd.edu, dataset identifier: MSV000085676). Figure 2 outlines these species and depicts their differing chemical characteristics and abundance ranges in serum.

Among common proteoforms, 3 apoAI backbone sequences were observed with different lengths: 1 containing the 6-residue propeptide of apoAI ("ProApoA-I"); the "canonical" backbone (residues D1-Q243); and a "truncated" backbone, missing the C-terminal Q243 residue. Furthermore, glycation of

Table 2. Characteristics of the CARDIA Participant Sample at the Year 20 Exam Stratified by HDL-C/CAC Groups

Group	CAC >0, HDL-C >60	CAC >0, 40 ≤ HDL-C ≤60	CAC >0, HDL-C <40	CAC=0, HDL-C >60	CAC=0, 40 ≤ HDL-C ≤60	CAC=0, HDL-C <40
n	25	25	25	25	25	25
Women, %	72	36	12	80	52	24
Age, y	47 (3)	46 (4)	47 (3)	47 (3)	44 (4)	44 (4)
Education, y	15 (3)	15 (3)	16 (2)	15 (2)	14 (3)	15 (2)
BMI, kg/m^2	29 (6)	31 (7)	33 (5)	27 (5)	29 (8)	31 (4)
Height, cm	171 (11)	171 (9)	178 (9)	168 (9)	171 (10)	173 (10)
Weight, lbs	184 (40)	201 (43)	228 (38)	168 (34)	185 (42)	205 (33)
Waist circumference, cm	89 (13)	97 (13)	108 (13)	85 (12)	91 (14)	99 (11)
Physical activity	261 (198)	331 (253)	337 (222)	337 (273)	365 (311)	390 (273)
Alcohol, mL/dL	17 (18)	12 (12)	6 (14)	16 (22)	22 (68)	6 (15)
Current smoking, %	32	16	8	16	12	24
SBP, mm Hg	120 (16)	119 (17)	117 (12)	108 (11)	117 (11)	115 (12)
DBP, mm Hg	74 (11)	73 (13)	73 (8)	68 (8)	71 (8)	71 (9)
Blood pressure-lowering medication, %	20	24	36	24	24	16
Total cholesterol, mg/dL	197 (41)	191 (37)	178 (36)	194 (29)	186 (54)	181 (32)
HDL-C, mg/dL	79 (23)	49 (6)	34 (4)	73 (10)	50 (5)	35 (3)
LDL-C, mg/dL	102 (34)	119 (37)	107 (31)	103 (30)	115 (50)	110 (29)
Triglycerides, mg/dL	82 (33)	114 (48)	194 (142)	92 (43)	107 (51)	177 (60)
Cholesterol-lowering medication, %	16	28	36	0	12	8
Diabetes mellitus medication, %	4	4	16	4	12	12
Fasting glucose, mg/dL	96 (32)	99 (20)	120 (54)	92 (11)	97 (14)	102 (14)
Diabetes mellitus, %	4	4	20	4	4	8
Total CAC, AU	116 (241)	70 (107)	156 (330)	0 (0)	0 (0)	0 (0)
HDL efflux	1.3 (0.2)	1.1 (0.2)	1 (0.2)	1.2 (0.2)	1.1 (0.2)	1.1 (0.2)
ABCA1-DEP. efflux	0.4 (0.2)	0.4 (0.2)	0.4 (0.2)	0.4 (0.2)	0.4 (0.2)	0.4 (0.2)
CRP	1.1 (1.7)	3.4 (4.8)	2.0 (2.3)	2.2 (2.3)	1.9 (2.6)	2.8 (3.1

Numbers represent either percent, for binary characteristics, or mean (SD), for continuous phenotype. ABCA1-DEP. efflux indicates ABCA1-dependent high-density lipoprotein cholesterol efflux; BP, blood pressure; CAC, coronary artery calcium; CARDIA, Coronary Artery Risk Development in Young Adults; CRP, C-reactive protein; DBP, diastolic blood pressure; LDL-C, low-density lipoprotein cholesterol; HDL, high-density lipoprotein; HDL-C, high-density lipoprotein cholesterol; Rx, "under medication for"; and SBP, systolic blood pressure.

residue K133 by a hexose was observed, along with 3 levels of methionine sulphoxidation (1, 2 and 3 oxidized methionine residues), all on the canonical apoAI backbone. Lastly, 7 of the 15 apoAI species observed were covalent additions of fatty acids (acylations) to the lysine 88 residue (K88acylApoAI) happening both on the canonical and the truncated backbone sequences. Consistently, these species eluted from reversed-phase liquid chromatography at a later retention time than the canonical proteoform because of their greater hydrophobicity (Figure S1). Further, while mass shifts of those proteoforms were consistent with modifications by fatty acids of 16, 18, 20, and 22 carbons, the combination of different unsaturated chains was not directly inferable from intact mass data, because of m/z overlap of apoAI ions differing by 2 hydrogens (1 unsaturation), as depicted in Figure S2.

These species were therefore treated as "proteoform groups" based on the number of carbons of the fatty-acid modification for purposes of proteoform quantification (vide infra).

In-depth characterization of these acylated forms by ion fragmentation during tandem mass spectrometry allowed for further insight on the nature of apoAI acylation (Figure S2). Notably, fragmentation patterns were consistent with acylations at K88 by 16:0, 18:0, 18:1, 18:2, 18:3, 20:4, 20:5, and 22:6 fatty acids. Of these, only palmitoylation (16:0) of apoAI had been described before our group's targeted top-down analyses of AapoAI.[12,13] Interestingly, these chain lengths and unsaturation states are consistent with the common types of fatty acids found in HDL particles.[21] Furthermore, the relative MS intensity of each acylated proteoform group (16C, 18C, 20C, and 22C) is also in

Proteoform or Proteoform Group	PFR ID*	Mass(es) (Da)	Side-chain PTMs /Backbone	%Abundance Ranges	Secondary Structure Representation
Canonical	4662605	28061.47	/Canonical (D1-Q243)	50-90%	M86 ... Q243
ProApoA-I	4662606	28943.90	/Propeptide-containing (R(-6)-Q243)	0.5-2%	... Q243
Truncation	4662607	27933.42	/Truncated (D1-T242)	3-20%	... T242
Oxidation	4662609	28077.47	Sulphoxidation at M86/ Canonical	4-27%	M86 - Sulphoxidation ... Q243
Di-Oxidations	---	28093.46	Sulphoxidation at M86 and M112 or M148/ Canonical	0.5-8%	M86 - Sulphoxidation / M112 - Sulphoxidation ... Q243
Tri-Oxidation	4662610	28109.46	Sulphoxidation at M86, M112 and M148/ Canonical	0.2-2%	M86 - Sulphoxidation / M112 - Sulphoxidation / M148 - Sulphox. ... Q243
Glycation	4662611	28223.53	Schiff base reaction of a Hexose to K133/ Canonical	0.2-2%	K133 - Glycation ... Q243
Phosphoric Acid Adduct	---	28158.44	Non-covalent Phosphate Adduct/ Canonical	0-2%	$H_2PO_4^-$ ... Q243
16C Acylations + Truncation	4662612 4662613	28171.65 28169.63	16:0 or 16:1 fatty acids at K88/ Truncated	0-0.2%	K88 - 16C Acylation ... T242
18C Acylations + Truncation	4662614 4662615 4662616	28199.68 28197.66 28195.65	18:0, 18:1 or 18:2 fatty acids at K88/ Truncated	0-0.3%	K88 - 18C Acylation ... T242
20C Acylations + Truncation	4662617 4662618	28219.65 28217.63	20:4 or 20:5 fatty acids at K88/ Truncated	0-0.1%	K88 - 20C Acylation ... T242
16C Acylations	4662619 4662620	28299.70 28297.69	16:0 or 16:1 fatty acids at K88/ Canonical	0.1-0.6%	K88 - 16C Acylation ... Q243
18C Acylations	4662621 4662622 4662623	28327.74 28325.72 28323.70	18:0, 18:1 or 18:2 fatty acids at K88/ Canonical	0.2-1.1%	K88 - 18C Acylation ... Q243
20C Acylations	4662624 4662625	28347.70 28345.69	20:4 or 20:5 fatty acids at K88/ Canonical	0.1-0.4%	K88 - 20C Acylation ... Q243
DHA Acylation	4662627	28371.70	22:6 (DHA) fatty acid at K88/ Canonical	0-0.2%	K88 - 22:6 Acylation ... Q243

Allelic Backbones

Isoform	Mass (Da)	Backbone	%Abundance Range	Secondary Structure Representation
F71Y	28077.47	Substitution F71Y	49% (of [ApoA-I])	Y71 ... Q243

* PFR IDs are searchable at the top-down proteomics proteoform repository (http://atlas.topdownproteomics.org).

Figure 2. The proteoforms of apoAI [apolipoprotein AI].
Top: The 15 species targeted for characterization and quantification. Three protein backbones were observed, characterized here by their first and last amino-acid residues in the canonical apoAI sequence (eg, R(-6)-Q243). Each single-proteoform species was given a unique proteoform identifier. Upon characterization, several species were found to be mixtures of different proteoforms, either of the same or very similar mass. These included di-oxidations; 16-, 18-, and 20-carbon acylations on a canonical backbone and on a truncated backbone. Each of these similar-mass sets was called a "proteoform group" and quantified as a single species. Proteoform abundance ranges were calculated based on total proteoform intensity divided by the summed intensity of all proteoforms of apoAI. Bottom: allelic backbones. An F71Y substituted backbone was observed, also presenting a set of 15 proteoforms and groups, created by the same modifications as the wildtype. The abundance range refers to the ratio of the summed intensity of proteoforms containing the allelic backbone to the sum of all ApoAI proteoform intensities in the heterozygotic individual observed. ApoAI indicates apolipoprotein AI; and PTM, post-translational modifications.

close agreement with the relative quantity of each fatty acid in HDL,[21] as shown in A of Figure S2.

To inquire on the mechanism of acylation in vivo, we incubated canonical apoAI with palmitic acid at average serum concentrations and serum pH (7.4) for 48 hours (Figure S3). Notably, while signal of the canonical proteoform remained similar to the one observed in serum samples, where palmitoylated proteoforms could be detected, no signal for these forms was observed above 0.1% relative abundance (the detection limit for this study). Therefore, for these incubation conditions, apoAI acylation did not happen spontaneously at detectable levels.

Characterization of Allelic Variants of ApoAI

In the 2 serum samples of 1 of the individuals studied (in years 20 and 25 of the CARDIA study), an allelic variant of apoAI was observed (Figure S4). For this participant, 2 apoAI isoforms of roughly equal abundance (one of which was roughly 16 Da higher in mass) were observed and characterized by fragmentation. Intact mass and fragment mapping patterns were consistent with a F71Y polymorphism,[22] indicating the individual was likely a heterozygote for *APOAI*. Interestingly, all the combinations of PTMs that were observed on the wildtype isoform of apoAI were also observed for this isoform.

Truncation and Dimerization of ApoAII

We also characterized the spectrum of apoAII chemical variation in the serum samples studied. Figure S5 shows raw LC-MS data for apoAII. Nine proteoforms of apoAII could be observed and characterized, as outlined in Figure 3. The most abundant conformation of ApoAII observed was dimeric, with 2 chains linked by a disulfide bridge at C6. The 6 dimeric proteoforms observed were results of different levels of C-terminal truncation of the 2 chains, as characterized by fragmentation and consistent with previous observations of apoAII dimers.[19] Moreover, we observed 3 monomeric proteoforms of apoAII, also differing by their degree of backbone truncation. Notably, all monomeric forms were modified by cysteinylation at C6. On average, around 7% of the backbones of apoAII observed were cysteinylated, and thus monomeric.

Apolipoprotein Proteoform Profile of 150 Individuals

We analyzed the proteoform profile of 150 individuals, sampled at years 20 and 25 of the CARDIA study. Figure S6 depicts parameters of quality assurance for this proteoform quantification. Notably, total MS intensity was strongly correlated (R^2=0.97) with concentration of apoAI standards analyzed during LC-MS

quantification blocks. Moreover, the total MS intensity of apoAI was significantly correlated to participant HDL-C values, at R=0.62, a coefficient consistent with previous results, acquired with different apoAI quantification tools.[23,24]

Variation in the percent abundance of apoAI and -AII proteoforms across the 150 individuals is displayed in Figure 4. Canonical was the most abundant apoAI proteoform observed, while truncation and different levels of oxidation were the most common modified forms within individuals (Figure 4A). Furthermore, although acylated proteoforms were among the least abundant, ranging from 0.5% to 2% of the total MS intensity observed, this proteoform family was the one that accounted for most of the relative abundance variation across individuals (Figure S7). In contrast to apoAI, apoAII proteoform profiles varied more widely across individuals (Figure 4B). Notably, while dimeric proteoforms were consistently more abundant than monomeric ApoAII, no single proteoform was the most abundant overall.

Covariance analysis of quantitative proteoform data (Figure S8) allowed for the characterization of proteoform "families," which vary similarly across individuals. An unbiased factor analysis highlighted 3 strongly correlated clusters of proteoforms in ApoAI and also 3 in ApoAII. The 3 correlation factors corresponding to these groups explained, in both cases, >80% of proteoform variation in the individuals analyzed. Interestingly, the proteoforms pertaining to each covarying group contained similar types of PTMs, namely: acylations, oxidations, and truncation (for ApoAI), and single truncation, double truncation, and dimerization (for ApoAII). This suggests that proteoform covariation in these systems is mostly attributable to a small number of common underlying mechanisms involving certain PTMs. Moreover, truncated forms of ApoAI and ApoAII also significantly covaried, further suggesting that some PTM mechanisms might be shared across proteins.

Cross-Sectional Associations of Proteoforms to Cardiometabolic Phenotype

We compared proteoform levels to individual cardiometabolic characteristics. In order to make this analysis independent of the association of total protein concentration to phenotype, we used the percent contribution of each proteoform to the total MS intensity of the apolipoprotein (Figure S8). Figure 5 shows heatmaps of the correlation coefficients relative to the association of each proteoform percent abundance to each continuous phenotype.

Notably, ApoAI proteoform heatmaps showed similar associations within covarying proteoform families and distinct associations between families. Moreover, these associations were similar between the 2 years studied.
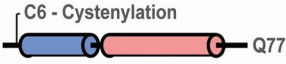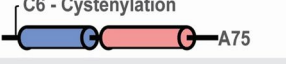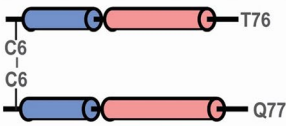
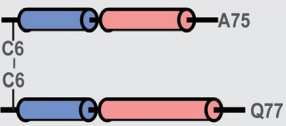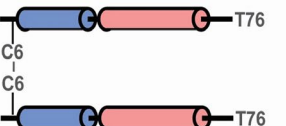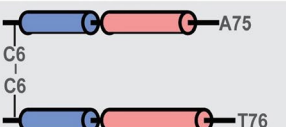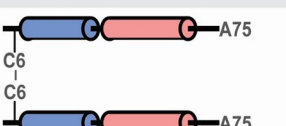| Proteoform	PFR ID	Mass (Da)	PTM or Backbone 1 /Backbone 2	%Abundance Range	Secondary Structure Representation
Cys/ATQ	4662651	8821.44	Cystenylation /Canonical (Q1-Q77)	2-12%	
Cys/AT	4662650	8693.38	Cystenylation /Singly Truncated (Q1-T76)	0-5%	
Cys/A	4662649	8592.34	Cystenylation /Doubly Truncated (Q1-A75)	0-3%	
ATQ/ATQ	4662654/ 4662654	17368.78	Canonical /Canonical	5-81%	
AT/ATQ	4662653/ 4662654	17240.73	Singly Truncated /Canonical	6-44%	
A/ATQ	4662652/ 4662654	17139.70	Doubly Truncated /Canonical	0-10%	
AT/AT	4662653/ 4662653	17112.69	Singly Truncated /Singly Truncated	0-22%	
A/AT	4662652/ 4662653	17011.64	Doubly Truncated /Singly Truncated	0-17%	
A/A	4662652/ 4662652	16910.60	Doubly Truncated /Singly Truncated	0-12%	

Figure 3. Proteoforms of ApoAII [apolipoprotein II].
Top: Three monomers of apoAII. All monomeric forms observed were cystenylated (modified by disulfide bridge with a free cysteine) at C6. As depicted on the left, 3 backbones were observed, either intact, doubly, or singly truncated. These are characterized here by their first and last amino-acid residues in the canonical ApoAII sequence (eg, Q1-A75). Monomers of apoAII were each given a unique proteoform identifier. Bottom: 6 dimers of ApoAII. Dimers were 2-way combinations of the backbones observed in the monomers, linked by disulfide bridge at C6. Proteoforms are named based on the last residues of their backbone and either cystenylation for monomers, or the last residues of the second backbone, for dimers (eg, AT/ATQ). Proteoform abundance ranges were calculated based on total proteoform intensity divided by the summed intensity of all proteoforms of ApoAII. PTM indicates post-translational modifications.

For instance, in both years 20 and 25 of CARDIA, hierarchical analysis clustered K88acylApoAI proteoforms together, and these forms showed the highest positive associations with HDL-C and HDL efflux. These acylated forms also showed negative associations with BMI and waist circumference. Furthermore, a similar pattern, albeit with overall weaker associations and effect size, was observed with the non-acylated truncated proteoform, which clustered with other truncated forms. These proteoforms also were negatively associated with common carotid thickness in both years. Conversely, the canonical proteoform showed

an overall opposite association pattern, with significant negative association to HDL-C and HDL efflux and positive associations to waist circumference, weight, and common carotid thickness. Mono-oxidized ApoAI, as well as the dioxidized form in year 25, showed a significant positive association with HDL-C, and, for year 20, mono-oxidation showed a positive association with C-reactive protein. Finally, as expected, we also observed a consistent and significant positive association between glycated ApoAI and serum glucose.

ApoAII proteoform heatmaps also showed distinct associations across the different proteoforms. In both
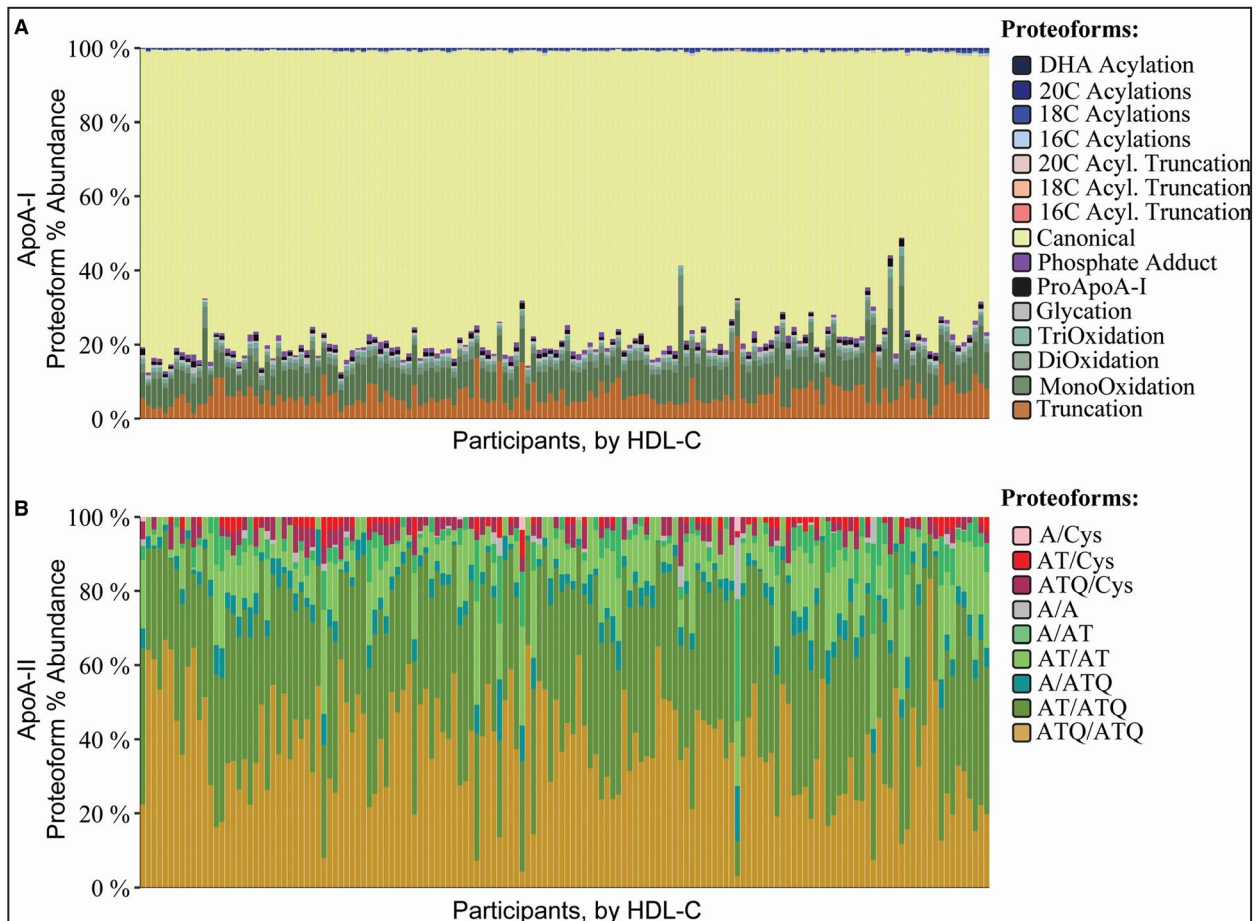
Figure 4. Proteoform profile of 150 individuals.
Colors depict the average contribution of each proteoform to the total intensity of ApoAI (apolipoprotein AI) in each individual, as rank-ordered by their high-density lipoprotein cholesterol. Values are shown for CARDIA year 20 samples only. **A**, ApoAI proteoforms. "Acyl." indicates acylations. **B**, ApoAII proteoforms. ApoAI indicates apolipoprotein AI; ApoAII, apolipoprotein II; and HDL-C, high-density lipoprotein cholesterol.

years 20 and 25 of CARDIA, the dimer of singly truncated chains (AT/AT) of ApoAII had the strongest positive association to HDL-C and HDL efflux. Concurrently, the fully non-truncated monomer (ATQ/Cys) and dimer (ATQ/ATQ) were clustered together, showing significant positive associations with markers of obesity (BMI, weight or waist circumference), the dimer also showed a negative association with HDL-C and HDL efflux. Finally, while most other forms showed no significant associations with phenotype, the overall clustering pattern suggests that proteoform-to-phenotype association is dependent both on number of truncations (singly truncated proteoforms were more strongly associated with higher HDL-C and lower obesity indices) and the number of chains present in the molecule (only dimers were associated with higher HDL-C and lower obesity indices).

For binary characteristics, a separate analysis was made (Figure S9). Notably, only one association was significant and consistent between both years of study: a higher level of glycated ApoAI was associated with diabetes mellitus.

Select proteoform-to-characteristic association data (R^2, beta coefficients and confidence parameters), including motif-based aggregates of proteoform intensity, are shown in Table S1. Most noteworthy, canonical abundance variation explained (based on R^2 values) 20% and 10% of the variation in HDL-C and HDL efflux in this dataset, while aggregate acylations explained 31% and 12%, respectively. However, associations with cardiometabolic indices varied in directionality: one positive standard deviation in canonical abundance was associated with −9.1 mg/dL of HDL-C and +3.9 cm of waist circumference, while beta coefficients were +11.4 mg/dL and −5.9 cm, respectively for the same characteristics and K88acylApoAI.

Longitudinal Analysis of Change in Proteoform Abundance to Change in Phenotype

We analyzed intra-individual changes in proteoform profile between years 20 and 25 of the CARDIA study and
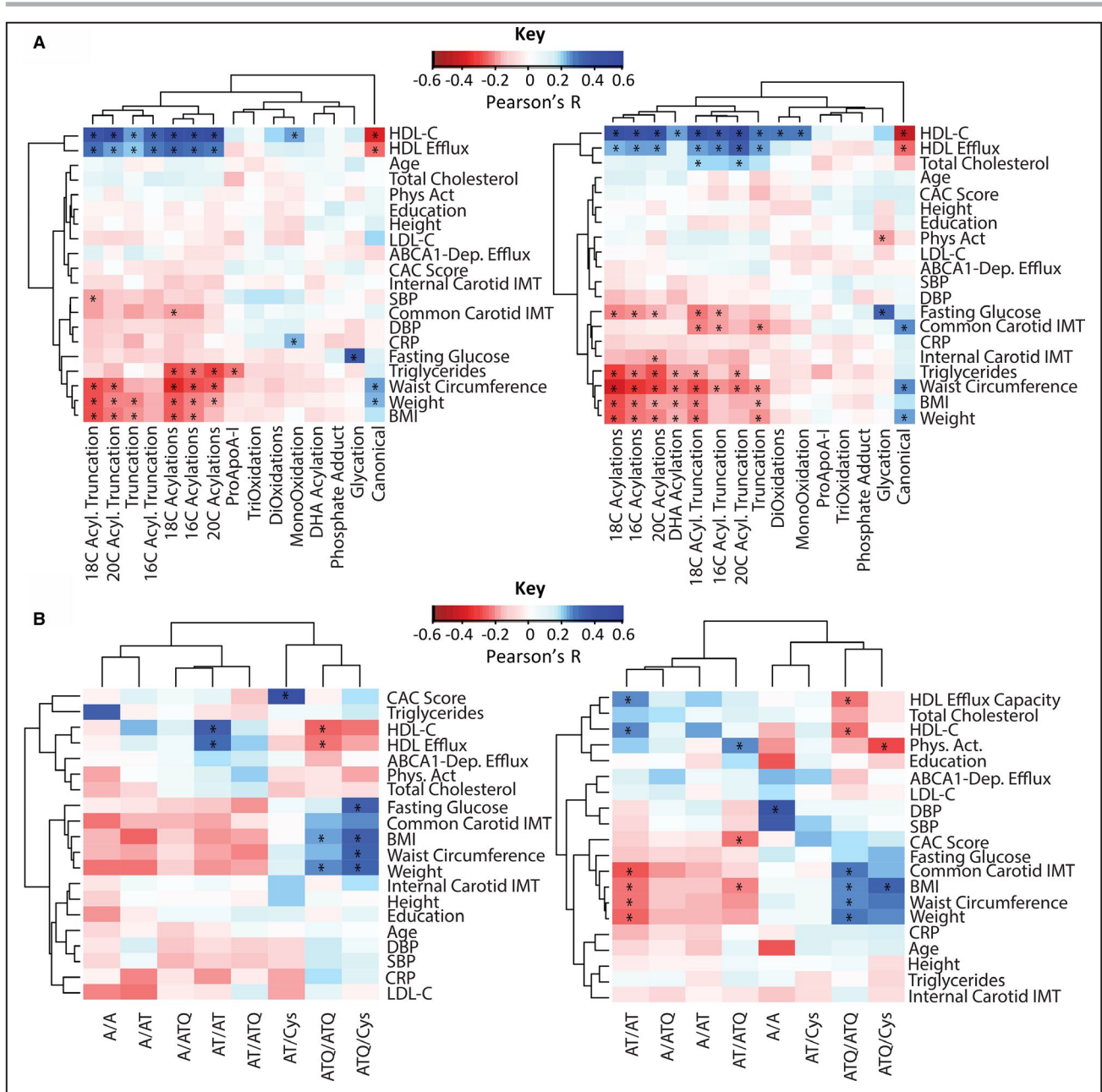
Figure 5. Correlation coefficient and significance of association of proteoforms percent abundance to continuous phenotype.

A, ApoAI proteoform associations. **B**, ApoAII proteoform associations. Left: CARDIA year 20. Right: CARDIA year 25. Abundances of each proteoform were compared with continuous phenotype in the 150 individuals studied. Clustering of both proteoforms and characteristics was unbiased. A Pearson *r* was generated for each correlation observed as well as a correlation *P* value. Colors show the strength and sign of each association. Statistical significance, symbolized by an asterisk, was asserted at 5% false discovery rate. ABCA1-Dep. efflux indicates ABCA1-dependent high-density lipoprotein cholesterol efflux; Acyl, acylations; BMI, body mass index; CAC, coronary artery calcium; CRP, C-reactive protein; DBP, diastolic blood pressure; HDL-C, high-density lipoprotein cholesterol; IMT, intima-media thickness; LDL-C, low-density lipoprotein cholesterol; Phys Act, physical activity; and SBP, systolic blood pressure.

compared them to change in participant characteristics over time. Figure 6 shows a heatmap of the correlation coefficients observed in this analysis. Notably, several associations observed in the cross-sectional analysis were also observed longitudinally. For instance, changes in truncated proteoforms were positively associated with changes in HDL-C, while change in canonical abundance

was inversely associated with change in HDL-C. Change in acylated forms were negatively associated with changes in BMI and waist circumference, and glycation was positively associated with changes in serum glucose. No significant associations were observed between change in ApoAII proteoform abundance overtime and change in cardiometabolic characteristics.

DISCUSSION

Summary of Results

In this analysis, we identified and quantified 15 distinct ApoAI proteoforms in 150 CARDIA participants for both CARDIA exam years 20 and 25, representing the largest cohort for proteoform measurement to date. Interestingly, top-down proteomics was able to identify gene-level variation as well. We observed, in one participant, a well-described and prevalent[22] allelic variant of ApoAI (F71Y), which also contained 15 distinct proteoforms, suggesting that there are no significant differences in ApoAI PTM physiology for this allele.

Moreover, associations with cardiometabolic traits appear to cluster ApoAI proteoforms in 4 motifs: canonical plus 3 proteoform families (truncations, oxidations,
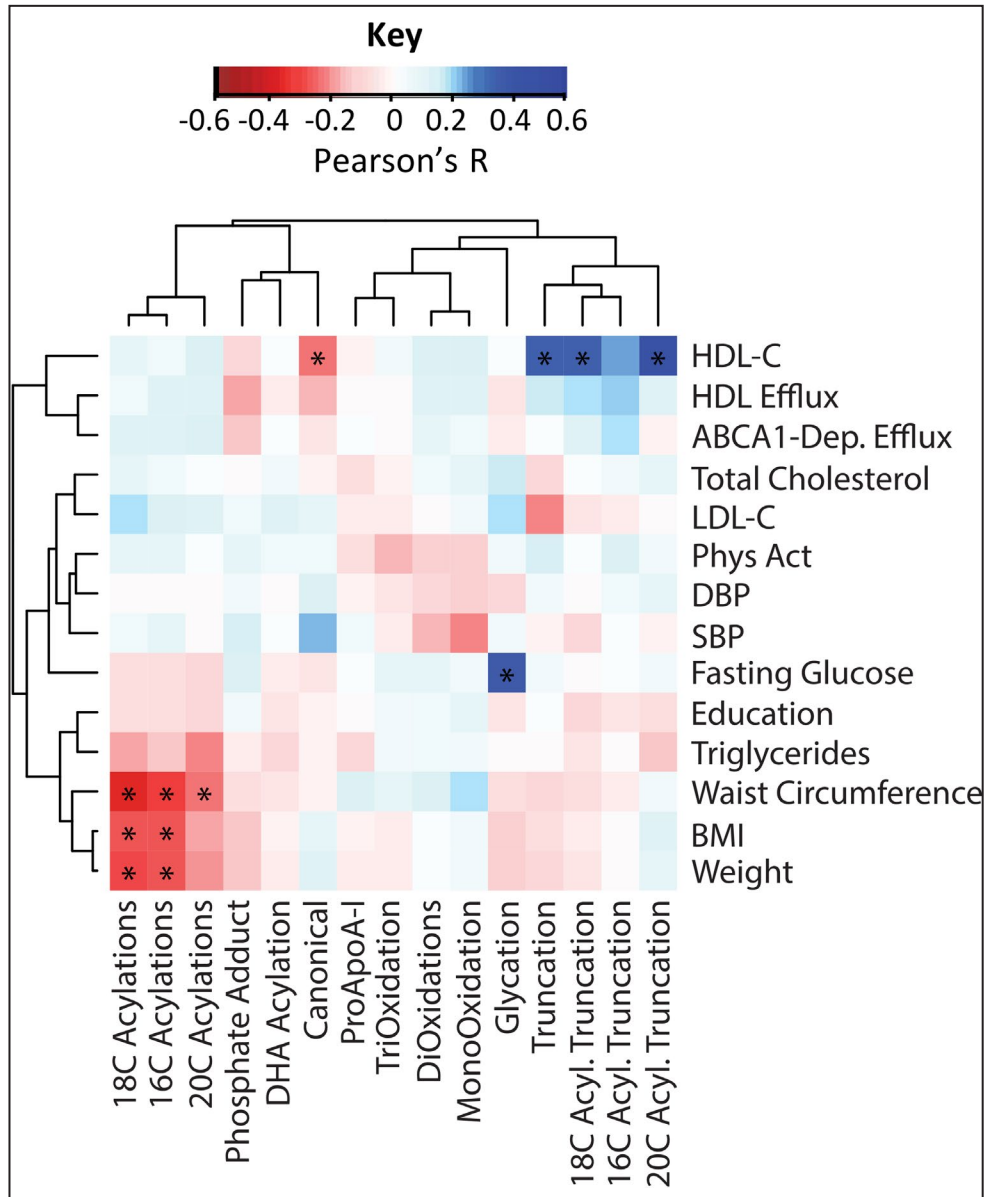
Figure 6. **Correlation coefficient and significance of association of intra-individual changes in ApoAI proteoform profile to changes in continuous phenotype.**
Changes in abundances of each proteoform were compared with changes in continuous phenotype within each of the 150 individuals studied. Clustering of both proteoforms and characteristics was unbiased. A Pearson r was generated for each correlation observed as well as a correlation P value. Colors show the strength and sign of each association. Statistical significance, symbolized by an asterisk, was asserted at 5% false discovery rate. ABCA1-Dep. efflux indicates ABCA1-dependent high-density lipoprotein cholesterol efflux; Acyl, acylations; BMI, body mass index; CAC, coronary artery calcium; CRP, C-reactive protein; DBP, diastolic blood pressure; HDL-C, high-density lipoprotein cholesterol; IMT, intima-media thickness; LDL-C, low-density lipoprotein cholesterol; Phys Act, physical activity; and SBP, systolic blood pressure.

and K88acylApoAI). We observed motif-specific associations between proteoform percent abundance and participant characteristics. For instance, percent canonical and K88acylApoAI proteoforms had significant associations with HDL-C, HDL efflux and markers of obesity, but the direction of association between these 2 proteoform motifs and indices of cardiometabolic health were opposite from each other (canonical was inversely associated with HDL-C and positively associated with waist circumference; vice-versa for K88acylApoAI).

Similarly, we observed motifs of proteoform covariation and associations with cardiometabolic phenotype in the 9 proteoforms of ApoAII. Importantly, while truncations of ApoAI and single truncations of ApoAII (both characterized by the loss of the C-terminal glutamine residue) covaried across individuals, common side-chain PTMs of ApoAI, such as acylations, had no analogous product in ApoAII. Furthermore, similar to ApoAI, the patterns of association of ApoAII proteoforms to cardiometabolic phenotype were largely proteoform-specific.

The proteoform-specific associations between ApoAI and AII and cardiometabolic health indices suggest that many proteoforms may be the cause or consequence of distinct (and possibly antagonistic) biological pathways involved in cardiometabolic health. The nature of this association to phenotype likely varies by proteoform. Figure 7A shows examples of hypothetical pathways through which metabolic, genetic and other biological factors may mediate the proteoform profile of ApoAI and AII. We posit that because specific proteoforms are the consequence of distinct biochemical pathways, which may be mediators or markers of phenotype, to the extent true, proteoform variation is more directly related to the biology underlying differences in phenotypic states than the observed variation in the aggregated concentration of proteins (Figure 7D through 7G).

Glycated ApoAI and Serum Glucose

Some proteoforms may result from enzymatic and non-enzymatic interactions with metabolites and thus they may serve as indices of metabolic dysregulation, which may not be detectable by measurement of total protein concentration. For example, analogous to hemoglobin A1c,[25] higher levels of glycated ApoAI are most likely the consequence of a well-described Schiff base reaction between serum glucose and serum proteins. Thus, similar to the hemoglobin A1c assay, glycated ApoAI appears to be a candidate marker of the aggregate exposure to serum glucose (Figure 7B). Interestingly the association between higher blood glucose levels and higher glycated ApoAI appears to be driven by participants with diabetes mellitus, suggesting, as for hemoglobin A1c, that glycation of ApoAI occurs in the setting of prolonged exposure to elevated blood glucose levels.[25] It follows then that the total intensity of glycated ApoAI is more strongly associated with serum glucose and diabetic status than total ApoAI (Figure 7E).

Acylation of ApoAI at Lysine 88 and HDL-C/HDL Efflux

We suspect acylation at K88 is an enzymatically mediated process because incubation of canonical ApoAI with palmitic acid did not spontaneously produce acylated proteoforms. Furthermore, acylated proteoforms of ApoAII were not observed, which argues that acylation is not a random occurrence on abundant HDL-associated proteins.

We also hypothesize that the abundance of K88acylApoAI may be a marker of the overall metabolic activity of ApoAI on HDL. ApoAI is known to exist in unbound and HDL-bound forms.[26–28] Interestingly, the fatty acids that are present on the acylated forms of ApoAI are the most prevalent fatty acids on HDL particles,[21] implying that the fatty acids added to ApoAI are derived from fatty acids imbedded in HDL particles (Figure 7C). We hypothesize that ApoAI acylation may be a consequence of promiscuous activity of HDL-maturation enzymes that use fatty acids as substrates (such as LCAT or PLTP, involved in cholesterol esterification and phospholipid formation, respectively). Thus, individuals with a larger percentage of K88acylApoAI

Figure 7. The differential association of proteoforms to phenotype.
A, Proteoform characterization integrates the effects of genetic variation and elements of the metabolome and possibly the exposome on the proteome. For example, ApoAI (apolipoprotein AI) proteoforms are the result of allelic variations of the *APOAI* gene, which is likely modified by enzymes coded for by other gene loci. **B**, Further, glycation likely occurs because of a well-described non-enzymatic Schiff-base reaction between serum glucose and proteins. **C**, Similarly, we hypothesize that acylation occurs on the high-density lipoprotein particle, suggesting that K88Acyl AI may be a marker of increased metabolic activity of ApoAI in high-density lipoprotein, which may explain the inverse associations between canonical ApoAI, acylated A-I, and markers of cardiometabolic health. **D**, Proteoform composition may vary significantly across biological states despite a smaller, or undetectable difference in total protein. Thus, measurement of total protein concentrations, for instance using standard ELISA assays, may fail to detect significant differences in proteoform abundance, which could modify the associations detected (as we demonstrate with ApoAI) and give insight into the biology that mediates phenotype. **E** through **G**, Examples of proteoform-specific associations with phenotype observed in this study. Crosses in (**E**) indicate participants with diabetes mellitus. For these panels absolute proteoform intensity was used, to compare with total apolipoprotein intensity. ApoAI indicates apolipoprotein AI; ApoAII, apolipoprotein II; HDL, high-density lipoprotein; and HDL-C, high-density lipoprotein cholesterol.
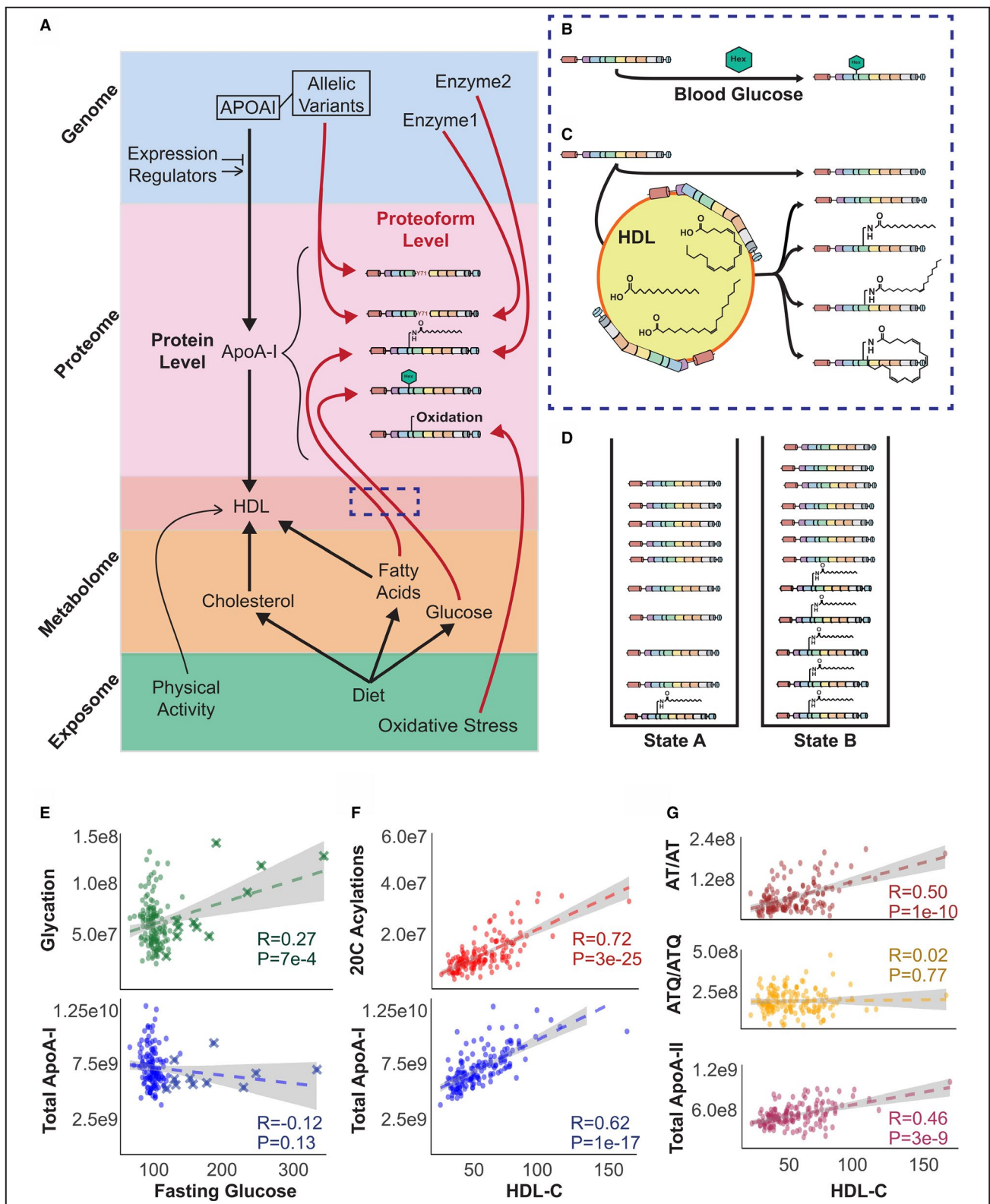
may not only have more HDL-active ApoAI, but also better HDL maturation (higher lecithin-cholesterol acyltransferase or phospholipid transfer protein activity) and consequently more reverse cholesterol transport and TG exchange with apoB particles, and potentially lower risk for developing advanced or unstable atherosclerosis. Conversely, a higher percentage of canonical ApoAI may represent a larger relative amount of free, non-metabolically active ApoAI, which may explain its inverse association with HDL-C.

Dimerization, Truncation of ApoAII and HDL-C

ApoAII is the second most prevalent protein on HDL particles. It is known to exist in monomeric, dimeric, singly- and doubly-truncated forms as we report in this study.[19,29,30] The role of ApoAII in HDL structure and function is less well defined than ApoAI. ApoAII is thought to serve as a competitive antagonist to ApoAI and thus modulate the functions of enzymes like lecithin-cholesterol acyltransferase and hepatic lipase that use ApoAI as a cofactor.[17,18] Further, limited data suggest that some allelic variants of ApoAII are associated with visceral adiposity and CHD.[30] This study is the first that provides a comprehensive assessment of ApoAII proteoform relative distributions and their associations with cardiometabolic phenotypes in humans. Given the prevalence of ApoAII in HDL particles, its previously described role in HDL structure and function, and the associations we report in this paper, it is possible that truncation of the terminal glutamine and dimerization of ApoAII may be involved with the activation of ApoAII's role in lipid metabolism. Moreover, cysteinylation of monomeric chains might regulate ApoAII dimerization and thus negatively affect its HDL-associated function. Therefore PTM-regulated differences in activity might underpin the stark difference in association to cardiometabolic indices observed between intensity of specific proteoforms and total ApoAII intensity (Figure 7G).

Implications for Lipid Biology

Our results add to a rapidly growing understanding of the HDL-associated proteome. Over 120 proteins are associated with HDL particles and the proteome varies significantly by HDL particle composition, size, number, and function.[16,31,32] The contribution of PTMs to the phenotypic and functional diversity of HDL particles in serum has not been well described and proteoform-level analyses can help fill this knowledge gap.

We believe these data have broader implications for both proteomics technology and lipid biology. For the latter, it is noteworthy that other apolipoproteins undergo post-translational modification, and these modifications likely have functional significance.[33–36] The findings we report in this analysis suggest that identification of apolipoprotein proteoforms may be a very useful way to gain novel insight into the complex biology of lipoprotein metabolism and its associations with states of health and disease. Furthermore, describing the differences in proteoform profiles across different tissues and disease states will provide important insights into the pathobiology of health and disease. Although we report ApoAI proteoforms in the serum compartment, there have been reports of a W72-oxidized proteoform of ApoAI in human atheroma,[15] which was not identified in this sample of human serum, suggesting that proteoform composition and the relative abundance of their PTMs may vary substantially by tissue compartment.

Strengths and Limitations

Using a methodology that we had previously employed in a small pilot sample from the Chicago Healthy Aging Study study,[12] we identified similar ApoAI proteoform motifs, characterized new ApoAII motifs, and we greatly expanded up on our understanding of their associations with cardiometabolic phenotypes. CARDIA allows for linkage of proteoform data to high-quality assessment of asymptomatic free-living individuals. Thus, we are able to examine the spectrum of proteoform diversity in the context of extensive participant phenotyping.

However, this analysis should be interpreted in the context of its limitations, as well. First, the dynamic range of current top-down proteomic technology does not allow for the identification of ApoAI proteoforms present in <0.1% relative abundance. Second, top-down proteomics requires label-free quantification, which provides relative quantification of proteoforms. However, MS intensity was highly correlated with absolute ApoAI concentration standards and total ApoAI MS intensity correlated with HDL-C at correlation coefficients previously reported in the literature using standard, high-fidelity assays. This suggests that the rank-order and scale of ApoAI proteoform quantification by top-down mass spectrometry is accurate. Third, the effect sizes reported are modest, but given the random error intrinsic to proteoform quantification the true biological associations are likely much stronger than what we report. Fourth, some of the proteoforms could be byproducts of sample preparation or electrospray ionization. In fact, we suspect this is the case of at least some of the oxidized proteoforms that we identified. However, the significant differences in the abundance of oxidized forms between individuals despite the same preparation technique and randomized injection, and a significant intra-individual correlation of oxidation abundance between the 2 years analyzed suggest that a biological pathway mediates these differences, not electrospray ionization or sample preparation.

CONCLUSIONS

We report the correlation of 21 total proteoforms of ApoAI and ApoAII across 150 people and show their quantitative analysis down to <1% relative abundance. It is noteworthy that, while each proteoform is an independent mass spectrometric measurement, covariance of proteoforms was consistent within types of PTMs (motifs), the associations between these proteoforms and indices of cardiometabolic health were

specific to and consistent within proteoform motifs and these associations portrayed externally consistent metabolic profiles. The most interesting example were acylations of ApoAI, which strongly covaried and were significantly and positively associated with indices of cardiometabolic health, such as a low waist circumference and high HDL-C, whereas the canonical form of ApoAI had opposite directions of association with the same cardiometabolic characteristics. Moreover, the discordance in associations between total protein concentration and specific proteoform intensities with cardiometabolic characteristic(s) highlights the importance of being precise about the biochemistry of proteins for biomarker research. Our data provide the largest-scale identification and associations between apolipoprotein proteoforms and human phenotypes. As such, the work serves as an example of using proteoform-resolved measurements to improve efficiency in gaining insight into the complex pathways through which proteins mediate health and disease.

ARTICLE INFORMATION

Affiliations

Department of Medicine (Cardiology) and Department of Preventive Medicine (J.T.W., D.M.L.); Department of Chemistry, Chemistry of Life Processes Institute and Proteomics Center of Excellence (H.S.S., P.D.C., R.L., P.F.D., N.L.K.); and Department of Medicine (Urology) (J.R., C.S.T.), Northwestern University, Chicago, IL; Department of Molecular Biology, University of Oklahoma, Norman, OK (L.F.); Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN (D.J.); and Mike and Valeria Rosenbloom Centre for Cardiovascular Prevention, Department of Medicine, McGill University Health Centre, Montreal, Quebec, Canada (A.S.).

Acknowledgments

Sources of Funding

Disclosures

Supplementary Material

Data S1
Table S1
Figures S1–S9
References 12, 37–42

REFERENCES

1. Andaç M, Denizli A. Affinity-recognition-based polymeric cryogels for protein depletion studies. *RSC Adv*. 2014;4:31130–31141. DOI: 10.1039/C4RA02655A.

2. Arciello A, Piccoli R, Monti DM. Apolipoprotein A-I: the dual face of a protein. *FEBS Lett*. 2016;590:4171–4179. DOI: 10.1002/1873-3468.12468.

3. Davidson WS, Thompson TB. The structure of apolipoprotein A-I in high density lipoproteins. *J Biol Chem*. 2007;282:22249–22253. DOI: 10.1074/jbc.R700014200.

4. Toth PP, Barter PJ, Rosenson RS, Boden WE, Chapman MJ, Cuchel M, D'Agostino RB, Davidson MH, Davidson WS, Heinecke JW, et al. High-density lipoproteins: a consensus statement from the National Lipid Association. *J Clin Lipidol*. 2013;7:484–525. DOI: 10.1016/j.jacl.2013.08.001.

5. McQueen MJ, Hawken S, Wang X, Ounpuu S, Sniderman A, Probstfield J, Steyn K, Sanderson JE, Hasani M, Volkova E, et al. Lipids, lipoproteins, and apolipoproteins as risk markers of myocardial infarction in 52 countries (the INTERHEART study): a case-control study. *Lancet*. 2008;372:224–233. DOI: 10.1016/S0140-6736(08)61076-4.

6. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, Hindy G, Hólm H, Ding EL, Johnson T, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet*. 2012;380:572–580. DOI: 10.1016/S0140-6736(12)60312-2.

7. Karjalainen MK, Holmes MV, Wang Q, Anufrieva O, Kähönen M, Lehtimäki T, Havulinna AS, Kristiansson K, Salomaa V, Perola M, et al. Apolipoprotein A-I concentrations and risk of coronary artery disease: a Mendelian randomization study. *Atherosclerosis*. 2020;299:56–63. DOI: 10.1016/j.atherosclerosis.2020.02.002.

8. Saleheen D, Scott R, Javad S, Zhao W, Rodrigues A, Picataggi A, Lukmanova D, Mucksavage ML, Luben R, Billheimer J, et al. Association of HDL cholesterol efflux capacity with incident coronary heart disease events: a prospective case-control study. *Lancet Diabetes Endocrinol*. 2015;3:507–513. DOI: 10.1016/S2213-8587(15)00126-6.

9. Rohatgi A, Khera A, Berry JD, Givens EG, Ayers CR, Wedin KE, Neeland IJ, Yuhanna IS, Rader DR, de Lemos JA, et al. HDL cholesterol efflux capacity and incident cardiovascular events. *N Engl J Med*. 2014;371:2383–2393. DOI: 10.1056/NEJMoa1409065.

10. Smith LM, Kelleher NL. Proteoform: a single term describing protein complexity. *Nat Methods*. 2013;10:186–187. DOI: 10.1038/nmeth.2369.

11. Mazur MT, Cardasis HL, Spellman DS, Liaw A, Yates NA, Hendrickson RC. Quantitative analysis of intact apolipoproteins in human HDL by top-down differential mass spectrometry. *Proc Natl Acad Sci USA*. 2010;107:7728–7733. DOI: 10.1073/pnas.0910776107.

12. Seckler HDS, Fornelli L, Mutharasan RK, Thaxton CS, Fellers R, Daviglus M, Sniderman A, Rader D, Kelleher NL, Lloyd-Jones DM, et al. A targeted, differential top-down proteomic methodology for comparison of ApoA-I proteoforms in individuals with high and low HDL efflux capacity. *J Proteome Res*. 2018;17:2156–2164. DOI: 10.1021/acs.jproteome.8b00100.

13. Hoeg JM, Meng MS, Ronan R, Fairwell T, Brewer HB Jr. Human apolipoprotein A-I. Post-translational modification by fatty acid acylation. *J Biol Chem*. 1986;261:3911–3914. DOI: 10.1016/S0021-9258(17)35598-9.

14. Shao B, Bergt C, Fu X, Green P, Voss JC, Oda MN, Oram JF, Heinecke JW. Tyrosine 192 in apolipoprotein A-I is the major site of nitration and chlorination by myeloperoxidase, but only chlorination markedly impairs ABCA1-dependent cholesterol transport. *J Biol Chem*. 2005;280:5983–5993. DOI: 10.1074/jbc.M411484200.

15. Huang Y, DiDonato JA, Levison BS, Schmitt D, Li L, Wu Y, Buffa J, Kim T, Gerstenecker GS, Gu X, et al. An abundant dysfunctional apolipoprotein A1 in human atheroma. *Nat Med*. 2014;20:193–203. DOI: 10.1038/nm.3459.

16. Wilkins JT, Seckler HS. HDL modification: recent developments and their relevance to atherosclerotic cardiovascular disease. *Curr Opin Lipidol*. 2019;30:24–29. DOI: 10.1097/MOL.0000000000000571.

17. Castellani LW, Nguyen CN, Charugundla S, Weinstein MM, Doan CX, Blaner WS, Wongsiriroj N, Lusis AJ. Apolipoprotein AII is a regulator of very low density lipoprotein metabolism and insulin resistance. *J Biol Chem*. 2008;283:11633–11644. DOI: 10.1074/jbc.M708995200.

18. Melchior JT, Street SE, Andraski AB, Furtado JD, Sacks FM, Shute RL, Greve EI, Swertfeger DK, Li H, Shah AS, et al. Apolipoprotein A-II

alters the proteome of human lipoproteins and enhances cholesterol efflux from ABCA1. *J Lipid Res*. 2017;58:1374–1385. DOI: 10.1194/jlr. M075382.

19. Honda K, Katzke VA, Hüsing A, Okaya S, Shoji H, Onidani K, Olsen A, Tjønneland A, Overvad K, Weiderpass E, et al. CA19-9 and apolipo-protein-A2 isoforms as detection markers for pancreatic cancer: a prospective evaluation. *Int J Cancer*. 2019;144:1877–1887. DOI: 10.1002/ijc.31900.

20. Friedman GD, Cutter GR, Donahue RP, Hughes GH, Hulley SB, Jacobs DR Jr, Liu K, Savage PJ. CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J Clin Epidemiol*. 1988;41:1105–1116. DOI: 10.1016/0895-4356(88)90080-7.

21. Noori M, Darabi M, Rahimipour A, Rahbani M, Abadi NA, Darabi M, Ghatrehsamani K. Fatty acid composition of HDL phospholipids and coronary artery disease. *J Clin Lipidol*. 2009;3:39–44. DOI: 10.1016/j.jacl.2008.11.010.

22. Haase CL, Frikke-Schmidt R, Nordestgaard BG, Tybjærg-Hansen A. Population-based resequencing of APOA1 in 10,330 individuals: spectrum of genetic variation, phenotype, and comparison with extreme phenotype approach. *PLoS Genet*. 2012;8:e1003063. DOI: 10.1371/journal.pgen.1003063.

23. Wu X, Yu Z, Su W, Isquith DA, Neradilek MB, Lu N, Gu F, Li H, Zhao XQ. Low levels of ApoA1 improve risk prediction of type 2 diabetes mellitus. *J Clin Lipidol*. 2017;11:362–368. DOI: 10.1016/j.jacl.2017.01.009.

24. Ma M-Z, Yuan S-Q, Chen Y-M, Zhou Z-W. Preoperative apolipoprotein B/apolipoprotein A1 ratio: a novel prognostic factor for gastric cancer. *Onco Targets Ther*. 2018;11:2169–2176.

25. Bunn HF, Haney DN, Kamin S, Gabbay KH, Gallop PM. The biosynthesis of human hemoglobin A1c. Slow glycosylation of hemoglobin in vivo. *J Clin Invest*. 1976;57:1652–1659. DOI: 10.1172/JCI108436.

26. Silva RA, Huang R, Morris J, Fang J, Gracheva EO, Ren G, Kontush A, Jerome WG, Rye KA, Davidson WS. Structure of apolipoprotein A-I in spherical high density lipoproteins of different sizes. *Proc Natl Acad Sci USA*. 2008;105:12176–12181. DOI: 10.1073/pnas.0803626105.

27. Panagotopulos SE, Horace EM, Maiorano JN, Davidson WS. Apolipoprotein A-I adopts a belt-like orientation in reconstituted high density lipoproteins. *J Biol Chem*. 2001;276:42965–42970. DOI: 10.1074/jbc.M106462200.

28. Melchior JT, Walker RG, Cooke AL, Morris J, Castleberry M, Thompson TB, Jones MK, Song HD, Rye K-A, Oda MN, et al. A consensus model of human apolipoprotein A-I in its monomeric and lipid-free state. *Nat Struct Mol Biol*. 2017;24:1093–1099. DOI: 10.1038/nsmb.3501.

29. Curtiss LK, Witztum JL. Plasma apolipoproteins AI, AII, B, CI, and E are glucosylated in hyperglycemic diabetic subjects. *Diabetes*. 1985;34:452–461. DOI: 10.2337/diab.34.5.452.

30. Xiao J, Zhang F, Wiltshire S, Hung J, Jennens M, Beilby JP, Thompson PL, McQuillan BM, McCaskie PA, Carter KW, et al. The apolipoprotein AII rs5082 variant is associated with reduced risk of coronary artery disease in an Australian male population. *Atherosclerosis*. 2008;199:333–339. DOI: 10.1016/j.atherosclerosis.2007.11.017.

31. Davidsson P, Hulthe J, Fagerberg B, Camejo G. Proteomics of apoli-poproteins and associated proteins from plasma high-density lipopro-teins. *Arterioscler Thromb Vasc Biol*. 2010;30:156–163. DOI: 10.1161/ATVBAHA.108.179317.

32. Jorge I, Burillo E, Mesa R, Baila-Rueda L, Moreno M, Trevisan-Herraz M, Silla-Castro JC, Camafeita E, Ortega-Muñoz M, Bonzon-Kulichenko E, et al. The human HDL proteome displays high inter-individual variability and is altered dynamically in response to angioplasty-induced atheroma plaque rupture. *J Proteomics*. 2014;106:61–73. DOI: 10.1016/j.jprot.2014.04.010.

33. Fievet C, Igau B, Bresson R, Drouin P, Fruchart JC. Non-enzymatic glycosylation of apolipoprotein A-I and its functional consequences. *Diabete Metab*. 1995;21:95–98.

34. Zhang Y, Sinaiko AR, Nelsestuen GL. Glycoproteins and glycosylation: apolipoprotein c3 glycoforms by top-down MALDI-TOF mass spectrometry. *Methods Mol Biol*. 2012;909:141–150.

35. Mancera-Páez O, Estrada-Orozco K, Mahecha M, Cruz F, Bonilla-Vargas K, Sandoval N, Guerrero E, Salcedo-Tacuma D, Melgarejo J, Vega E, et al. Differential methylation in APOE (Chr19; Exon four; from 44,909,188 to 44,909,373/hg38) and increased apolipoprotein E plasma levels in subjects with mild cognitive impairment. *Int J Mol Sci*. 2019;20:1394. DOI: 10.3390/ijms20061394.

36. Jian W, Edom RW, Wang D, Weng N, Zhang SW. Relative quantitation of glycoisoforms of intact apolipoprotein C3 in human plasma by liquid chromatography-high-resolution mass spectrometry. *Anal Chem*. 2013;85:2867–2874. DOI: 10.1021/ac3034757.

37. Wilkins JT, Li RC, Sniderman A, Chan C, Lloyd-Jones DM. Discordance between apolipoprotein B and LDL-cholesterol in young adults predicts coronary artery calcification: the CARDIA study. *J Am Coll Cardiol*. 2016;67:193–201. DOI: 10.1016/j.jacc.2015.10.055.

38. de la Llera-Moya M, Drazul-Schrader D, Asztalos BF, Cuchel M, Rader DJ, Rothblat GH. The ability to promote efflux via ABCA1 determines the capacity of serum specimens with similar high-density lipoprotein cholesterol to remove cholesterol from macrophages. *Arterioscler Thromb Vasc Biol*. 2010;30:796–801. DOI: 10.1161/ATVBAHA.109.199158.

39. Luthi AJ, Lyssenko NN, Quach D, McMahon KM, Millar JS, Vickers KC, Rader DJ, Phillips MC, Mirkin CA, Thaxton CS. Robust passive and active efflux of cellular cholesterol to a designer functional mimic of high density lipoprotein. *J Lipid Res*. 2015;56:972–985. DOI: 10.1194/jlr. M054635.

40. DeHart CJ, Fellers RT, Fornelli L, Kelleher NL, Thomas PM. Bioinformatics analysis of top-down mass spectrometry data with ProSight Lite. *Methods Mol Biol*. 2017;1558:381–394.

41. Ntai I, Kim K, Fellers RT, Skinner OS, Smith AD IV, Early BP, Savaryn JP, LeDuc RD, Thomas PM, Kelleher NL. Applying label-free quantitation to top down proteomics. *Anal Chem*. 2014;86:4961–4968. DOI: 10.1021/ac500395k.

42. Gorshkov MV, Fornelli L, Tsybin YO. Observation of ion coalescence in Orbitrap Fourier transform mass spectrometry. *Rapid Commun Mass Spectrom*. 2012;26:1711–1717. DOI: 10.1002/rcm.6289.

SUPPLEMENTAL MATERIAL

Data S1.

Supplemental Methods

Phenotypic Assessment

Race, sex, medication use, tobacco use, alcohol use and prevalent CHD status were assessed via self-report. Blood pressure was measured three times using an automatic oscillometer (OmROn) with the participant seated; the average of the second and third measurements was used. Height, weight, and waist circumference were measured using standard techniques. Blood was obtained via venipuncture of the antecubital vein from participants in the fasting state, and serum was sent for lipid sub-fraction analysis as well as fasting glucose. Samples were frozen at -80°C indefinitely.

CAC Measurement:

At year 20, 2 scans were obtained per participant. Using either a GE Lightspeed QX/I, Imatron C-150, or Siemens S4+ Volume Zoom axial 2.5-3 mm slices along the long axis of the heart were obtained. Images were transmitted electronically to the reading center for analysis[37].

HDL Efflux Measurement:

Cholesterol efflux experiments were carried out as described previously[38,39]. Briefly, J774 cells (murine macrophages) were plated into 24-well culture plates at a

density of 1.5×10^5 cells/ well. Following plating, cells were loaded with 2 µCi/ mL ^{3}H-cholesterol for 24 hours, with the small molecule inhibitor Sandoz 58-035 (2 µg/ mL) added to prevent cholesterol esterification. Post labeling, cells were treated with or without cyclic AMP (cAMP, 0.3 mM) overnight in order to upregulate ABCA1 expression. During this incubation, CARDIA serum samples were processed for use in the efflux assay. A total of 60 µL of serum was added to 24 µL of PEG-8000 (20% w/v dissolved in 200 mM glycine, pH = 7.4), incubated at room temperature for 20 minutes, then centrifuged at 10,000 rpm for 30 minutes at 4° C, in order to remove the LDL and VLDL components. Following centrifugation, 50.4 µL of supernatant (HDL containing fraction) was added to 1.7496 mL of MEM, containing 20 mM HEPES and 1% Penicillin-Streptomycin (MEM/HEPES/PS), as well as 2 µg/ mL Sandoz 58-035 and, for the cAMP efflux group, 0.03 mM cAMP. After overnight incubation with or without cAMP, cells were washed twice with MEM/HEPES/PS and 500 µL of media + serum samples were added to each well. Cells were incubated with the serum samples for 4 hours in a 37^oC, 5% CO_2 humidified incubator. After incubation, the culture media was filtered, added to 3 mL of Ultima Gold liquid scintillation fluid in 6 mL glass pony vials, and counted using a Perkin Elmer Tri-Carb 3100TR liquid scintillation counter. The resultant counts were standardized to the total cellular cholesterol levels prior to the addition of serum samples, and then against a pooled control sample.

Serum Preparation for Top-down Proteomics:

Frozen aliquots from banked serum were gently thawed and subjected to ApoB-depletion. To control for pre-analytical variation, all samples used for proteoform quantification were stored under the same conditions and preparation was performed with the same lot for each reagent, under controlled temperature and at the same time. For ApoB-depletion, four 10 µL aliquots of each serum sample were added each to 10 µL of LipoSep anti-ApoB immunoprecipitation resin (Sun Diagnostics, New Gloucester, ME). The ensuing centrifugation (performed as described in the manufacturer's protocol) depleted the serum of ApoB-bound particles such as LDL and VLDL. The then HDL-enriched supernatant was collected and a 1 µL aliquot was diluted in 120 µL of solvent A (5% acetonitrile and 0.2% formic acid in water), which was frozen until mass spectrometric analysis.

Liquid Chromatography/Mass Spectrometry (LC-MS):

Protein samples were subjected to reversed-phase liquid chromatography (RPLC) using an Ultimate 3000 LC system (Thermo Scientific, San Jose, CA). A 1 µL aliquot of each sample was loaded onto a trap column (20 mm, 150 µm inner diameter, i.d.) packed with PLRP-S resin (Agilent, Santa Clara, CA) for an initial wash. For separation of proteoforms, an in-house packed capillary PLRP-S column (200 mm, 75 µm i.d.) was used. Both columns were heated at 35° C. For proteoform quantification, the gradient consisted of a ramp of solvent B from 30 to 50% in 20 min., with a total run time of 35 min., including column wash (at 90% solvent B) and re-equilibration. For proteoform characterization, a longer, 1-hour gradient was used. Solvent A consisted of 5%

acetonitrile and 0.2% formic acid in water, while solvent B was composed of 5% water and 0.2% formic acid in acetonitrile. The outlet of the column was on-line coupled to a nanoelectrospray ionization source, to which a ~2 kV potential was applied for ionizing proteoforms for mass spectrometry (MS) analysis.

Quantitative mass spectrometry measurements were performed using an Orbitrap Elite (Thermo Scientific, Bremen, Germany), operating in "intact protein mode" (N_2 pressure in the HCD cell was set to 0.8 mTorr, and HCD trapping was turned on). The instrument operated in MS1-mode only (i.e., no fragmentation was applied). The instrument was cycling between a full scan (over a 500-2000 *m/z* window) and three selected-ion-monitoring (SIM) scans (over a 826.46 - 851.89 *m/z* window, to target the 34+, 33+ and 35+ charged states of ApoA-I proteoforms; a 878.09 - 906.57 *m/z* window, for the 32+, 31+ and 33+ charged states; and a 1178.27-1346.23 *m/z* window, to target ApoA-II monomers at the 7+ charged state). The applied resolving power was 120,000 (at 200 *m/z*), while the automatic gain control (AGC) target was set at 2.0E5 and 1.0E6 for broadband MS1 and SIM, respectively. All full MS and SIM scans were obtained by averaging 4 microscans.

Qualitative mass spectrometry (for proteoform characterization) was performed on an Orbitrap Fusion Lumos Tribrid (Thermo Scientific, San Jose, CA), operating in protein mode, with an HCD pressure setting of 0.2 mTorr. The instrument cycled between full MS and tandem MS scans. Ions were selected by 1 Da-wide quadrupolar isolation for fragmentation by either electron transfer dissociation (ETD), collision-assisted ETD (ETciD) or high-energy collision induced dissociation (HCD). The applied resolving power

was either 120,000 (for full MS) or 60,000 (for tandem MS), at 200 *m/z*. The automatic

gain control (AGC) target was set at 2.0E5 ($2x10^5$) for full MS and varied by proteoform

for tandem MS, ranging between 2.0E4-1.0E6 depending on proteoform abundance.

Maximum injection times for tandem MS were changed accordingly, to allow reaching the

AGC target. All full MS and tandem MS scans were obtained by averaging 4 microscans.

Proteoform Characterization from Top-down MS Data:

For ApoA-I proteoform characterization and PTM mapping, SIM scans of an initial

qualitative LC-MS run were manually analyzed to search for MS peaks of similar charge

distribution, mass and retention time of ApoA-I. All peaks that fit these parameters were

further analyzed by top-down fragmentation. Intact mass spectra were recorded and top-

down fragmentation data were acquired for each species. Then, ion masses

deconvoluted from the fragmentation and intact mass spectra were analyzed with the

freeware ProSight Lite (http://prosightlite.northwestern.edu/)[40] against the ApoA-I

sequence and modification variants found in the UniProt database (UniProt Accession

#P02647) and the biochemical and clinical literature on ApoA-I. All MS peaks that were

identified as species containing an ApoA-I backbone but for which the mass did not match

any previously known proteoforms were characterized with a classical top-down

proteomics approach for proteoform discovery. Briefly, the highly resolved intact mass

difference (Δm) of a proteoform to the canonical form was used to narrow the list of

possible chemical modifications or genetic variations to the proteoform and the presence

and location of each modification was tested by fragmentation coverage of the

modification site. The generated fragmentation maps were assigned P-scores for proteoform identification confidence. This same approach was then applied to ApoA-II.

Fragmentation data for every proteoform identified in this study, along with all graphical fragmentation maps used in this analysis containing identification confidence parameters (such as proteoform P-score) were uploaded to the MassIVE database (https//:massive.ucsd.edu), dataset identifier: MSV000085676. All proteoforms of ApoA-I detectable in this study were either matched to previously characterized proteoforms or *de novo* chemically characterized. Proteoforms were also given a permanent proteoform identifier number (PFR ID, or PFR#) and uploaded to the proteoform repository, maintained by the Consortium for Top Down proteomics (http://repository.topdownproteomics.org/).

Quantitative Comparison of Proteoforms

For quantification of the characterized proteoforms, six data acquisition blocks were set up, each a 10-day block in which 200 samples (four preparation replicates for each of 25 individuals for each of two years studied) were analyzed. Participants were randomly assigned to each block. All preparation replicates from years 20 and 25 exam cycles from the same individuals were analyzed in the same block, in order to reduce error in longitudinal analysis. In each block, data acquisition was set up as previously described for a quantitative top-down proteomics workflow[12]. Briefly, in order to mitigate the effects of measurement errors arising from sample preparation and instrumental

performance drift, samples were injected in a randomized fashion while avoiding clusters of samples from the same participant.

For proteoform abundance comparison, the generated SIM scans were searched for isotopic patterns of target proteoforms by a previously described fitter algorithm[41]. Briefly, a list of chemical formulas for the well-characterized ApoA-I proteoforms was used to calculate isotopic patterns for specific charge states of each form. Then, the centroided peaks of the observed spectrum above a minimum intensity (NL=100) were searched against the expected *m/z* of the two most abundant isotopomers calculated for the given chemical formulas. A window of 10 ppm was allowed for matches in order to account for mass accuracy drift. The full isotopic distribution for matching formulas was then adapted to the observed mass error and the spectra was searched for the remaining isotopomers with a 2 ppm allowance, to account for charge-space effects such as coalescence[42]. The relative intensities of the matched isotopic peaks were compared to the expected distribution in order to generate a fitting score. Fitting scores ranged from 0-1 and were calculated by a least squares model weighed positively towards higher intensities. Proteoform intensities were quantified for each SIM scan in which the relative chemical formula passed a fitting score threshold of 0.3. For each scan, background noise intensity was sampled and decreased from the signal of each proteoform, in order to reduce chemical noise effects in proteoform quantification. The final total intensity per LC-MS run was calculated for each proteoform using the area under the curve of the ensuing background-subtracted extracted-ion chromatogram.

To validate the association between ApoA-I MS intensity (as calculated by the aforementioned method) and ApoA-I abundance in the initial sample, 6 samples of known ApoA-I concentration (0.72, 1.44, 2.16, 2.88, 4.32 and 5.76 μg/μL) were prepared from a pure ApoA-I standard (MyBioSource, San Diego, CA, USA) and run with the same LC-MS method, in random order during each the 6 blocks of LC-MS runs. The proteoforms present in the standards were analyzed similarly to all quantitative samples in this study and the correlation between MS intensity and injected mass was described.

Data Analysis and Statistical Treatment

Quantitative MS intensities were normalized by total ion current of each LC-MS run to account for instrument drift. Normalized intensity values were then standardized between collection blocks to account for block-to-block variation: per-run and per-proteoform values were standardized so that the mean and standard deviations of each block were the overall mean and standard deviation values of the entire experiment, by the following method: standardized intensity = normalized intensity z-score (per block) x overall standard deviation + overall mean. Percent proteoform data (for the relative analysis) were then calculated for each run by dividing each proteoform intensity to the sum of the standardized intensities of all proteoforms. Either percent data or standardized intensity were then averaged per individual serum sample and compared to their characteristics by either linear regression for continuous data (accompanied with Pearson's r, beta coefficients and a Pearson's product-moment correlation p-score) or by t-test for binary data (accompanied with fold difference and p-scores). For sensitivity

analysis, linear models for microarray analysis (LIMMA) was also used for binary data comparisons, producing overall similar results (data not shown). Percent of total proteoform intensity was used in the primary analyses to normalize for differences in absolute ApoA-I and ApoA-II concentrations across individuals, allowing for greater insight into proteoform-specific effects. A Benjamini-Hochberg correction was used to multiple-test correct these scores and significance was asserted at a 5% false discovery rate (2-tailed type 1 error rate).

There were 2.2% missing values for proteoform abundance in the ApoA-I dataset and 17.4% for ApoA-II. These missing values arose from proteoforms there were not detected above the 0.3 threshold for the isotopic matching script. The monomer A/Cys of ApoA-II, commonly the lowest abundant proteoform, was responsible for 7.33% of ApoA-II missing values and was removed from the quantitative analysis. Missing values were removed from all quantitative analyses.

Results from linear regression (Pearson's R) or binary characteristic comparisons (fold difference between groups) were plotted onto a heatmap which clustered groups of proteoforms and characteristics with similar results, using the "heatmap.2" function contained in the "gplots" package for R. Briefly, "heatmap.2" relies on the "dist" function from package "stats" to calculate the Euclidian distance between the values input to the heatmap analysis and "hclust" (also from package "stats"), which performs the hierarchical clustering based on those distances. The method used for hierarchical clustering was the default for that function: complete-linkage clustering. The same

methodology was also used to plot a heatmap of correlation coefficients between relative abundances of proteoforms of ApoA-I and proteoforms of ApoA-II.

A factor analysis was also applied to proteoform abundance values, in order to identify groups of proteoforms that co-varied across individuals. To that end, the function "factanal" from R package "stats" was used, which performed maximum-likelihood factor analyses to a matrix of standardized proteoform abundances by participant. The rotation method "promax" was chosen for this analysis.

Table S1. Linear regression metrics for proteoforms and aggregated proteoform motifs.

		Year 20				Year 25			
ApoA-I		β	S.E. (β)	P-score	R²	β	S.E. (β)	P-score	R²
Truncation	BMI (kg/m²)	-1.6	0.5	1.5E-03	0.07	-1.4	0.5	9.7E-03	0.04
	C. CARO. IMT (mm)	-0.032	0.010	2.7E-03	0.06	-0.038	0.010	3.5E-04	0.09
	HDL EFFLUX	0.046	0.016	4.4E-03	0.05	0.050	0.015	1.3E-03	0.07
	HDL-C (mg/dl)	5.4	1.6	9.7E-04	0.07	6.1	1.5	7.8E-05	0.10
	WEIGHT (lbs)	-11.5	3.4	8.9E-04	0.07	-11.0	3.5	2.2E-03	0.06
	WAIST CIRC. (cm)	-3.1	1.2	8.0E-03	0.05	-3.9	1.3	2.2E-03	0.06
Canonical	HDL EFFLUX	-0.063	0.015	7.6E-05	0.10	-0.058	0.015	1.3E-04	0.09
	HDL-C (mg/dl)	-9.1	1.5	7.6E-09	0.20	-10.0	1.3	5.3E-12	0.28
	WEIGHT (lbs)	10.6	3.4	2.1E-03	0.06	11.5	3.5	1.3E-03	0.07
	WAIST CIRC. (cm)	3.9	1.2	8.1E-04	0.07	4.6	1.2	2.7E-04	0.09
Glycation	FA. GLUC. (mg/dl)	13.1	2.2	9.8E-09	0.20	9.4	2.0	4.6E-06	0.13
Acylation Truncations	BMI (kg/m²)	-2.0	0.5	7.2E-05	0.10	-1.6	0.5	3.2E-03	0.06
	C. CARO. IMT (mm)	-0.032	0.010	2.5E-03	0.06	-0.040	0.010	1.2E-04	0.10
	HDL Efflux	0.066	0.015	3.2E-05	0.11	0.060	0.015	7.7E-05	0.10
	HDL-C	9.6	1.5	1.1E-09	0.22	9.9	1.3	1.1E-11	0.27
	WEIGHT (lbs)	-13.6	3.3	7.3E-05	0.10	-10.5	3.5	3.5E-03	0.06
	WAIST CIRC. (cm)	-4.6	1.1	7.0E-05	0.10	-5.1	1.2	6.4E-05	0.10
Acylations	BMI (kg/m²)	-1.9	0.5	2.1E-04	0.09	-2.2	0.5	5.3E-05	0.11
	HDL EFFLUX	0.069	0.02	1.2E-05	0.12	0.051	0.015	1.0E-03	0.07
	HDL-C (mg/dl)	11.4	1.4	8.6E-14	0.31	9.9	1.3	1.1E-11	0.27
	TG (mg/dl)	-29.0	6.4	1.2E-05	0.12	-31.2	6.6	5.5E-06	0.13
	WEIGHT (lbs)	-13.0	3.3	1.5E-04	0.09	-14.5	3.4	4.4E-05	0.11
	WAIST CIRC. (cm)	-5.9	1.1	2.2E-07	0.17	-7.2	1.2	4.2E-09	0.21
ApoA-II									
ATQ/Cys	BMI (kg/m²)	2.0	0.7	3.0E-03	0.09	2.2	0.6	6.7E-04	0.11
AT/AT	HDL EFFLUX	0.046	0.015	2.8E-03	0.06	0.058	0.016	3.1E-04	0.08
	HDL-C (mg/dl)	4.7	1.5	2.3E-03	0.06	6.7	1.6	3.9E-05	0.11
ATQ/ATQ	BMI (kg/m²)	1.6	0.5	2.8E-03	0.06	1.5	0.5	3.5E-03	0.06

		β	S.E. (β)	p-value	R²	β	S.E. (β)	p-value	R²
	HDL EFFLUX	-0.046	0.015	2.9E-03	0.06	-0.050	0.016	2.1E-03	0.06
	HDL-C (mg/dl)	-4.6	1.5	2.8E-03	0.06	-5.7	1.6	5.3E-04	0.08
	WEIGHT (lbs)	11.7	3.5	1.1E-03	0.07	10.4	3.4	2.7E-03	0.06
Longitudinal									
Truncation	HDL-C (mg/dl)	3.8	0.9	5.3E-05	0.11				
Canonical	HDL-C (mg/dl)	-2.8	0.9	2.6E-03	0.06				
Glycation	FA. GLUC. (mg/dl)	8.7	2.0	2.1E-05	0.12				
Acyl. Trunc.	HDL-C (mg/dl)	4.4	0.9	2.4E-06	0.14				
Acylations	BMI (kg/m^2)	-0.8	0.2	3.1E-04	0.08				
	WEIGHT (lbs)	-5.2	1.3	1.4E-04	0.09				
	WAIST CIRC. (cm)	-2.2	0.4	9.5E-06	0.13				

Beta coefficients were standardized to the proteoform variation only, so each value represents the variation in metabolic characteristic associated with one standard deviation in proteoform percent abundance. Standard error of beta coefficients (S.E. (β)) are shown. All p-scores displayed were significant at a 5% Benjimani-Hochberg false discovery rate.

Figure S1. Mass spectra of the proteoforms of ApoA-I.
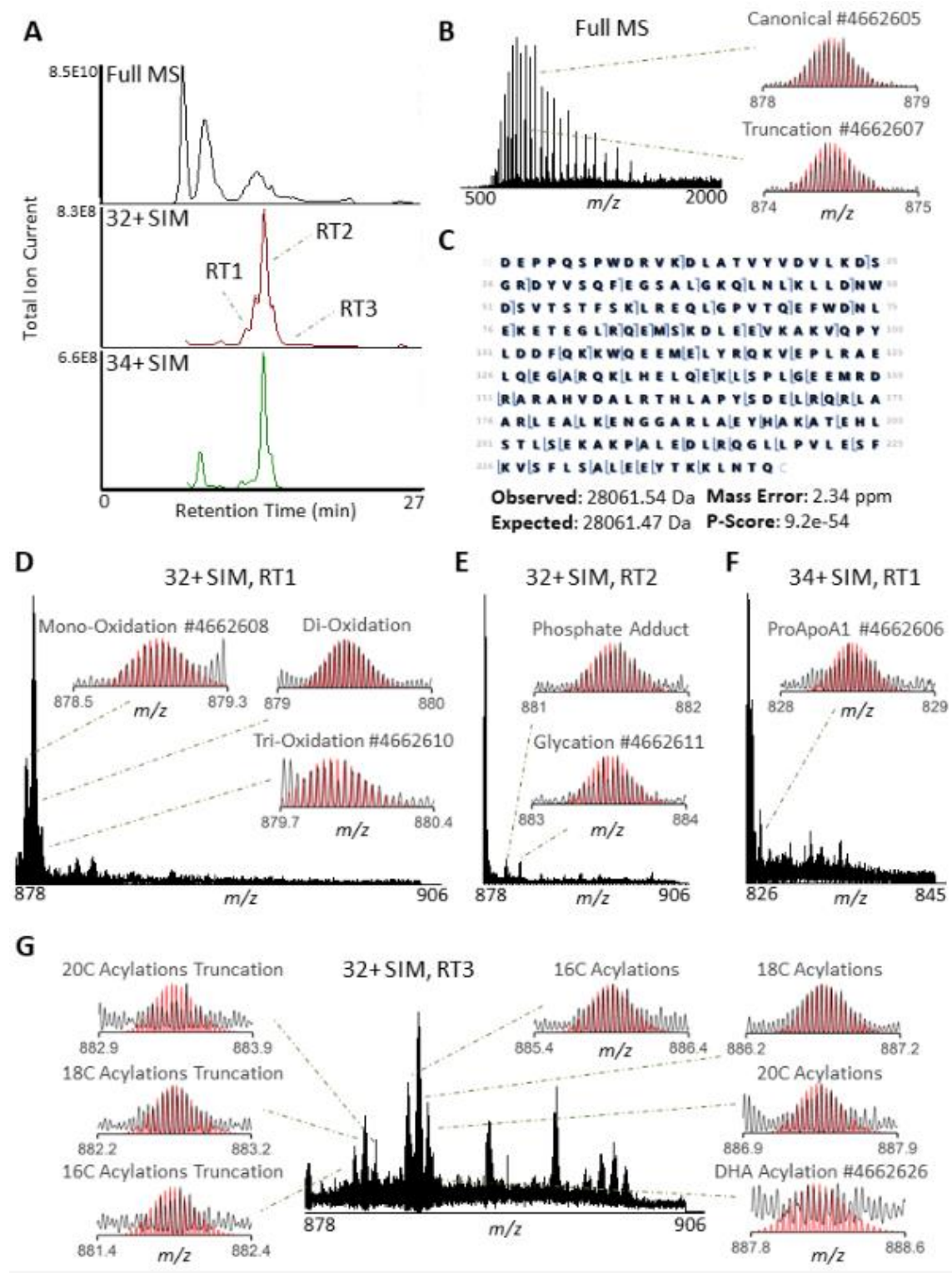
A: Example of a chromatogram of the 3 types of scans scheduled in the LC-MS method. In black, the full MS chromatogram of the run, a more encompassing *m/z* window (500-2000) was used in these scans, and multiple peaks, roughly corresponding to different proteins, can be observed. ApoA-I and its proteoforms were observable from retention time 11-19 min. In red and green, narrow, ApoA-I-specific SIM scans targeting the 32+ and 34+ charge states are shown, respectively. Different proteoforms were observed at 3 retention time windows (RT1, RT2 and RT3). **B:** Full MS scan at RT2. In this type of scan, multiple charged states of a mass corresponding to canonical ApoA-I were observed. The truncated proteoform was also observable in full MS. Insets show a comparison of the observed isotopic distribution, in black, of one charge state of either Canonical or Truncation and, in red, its theoretical expected distribution. **C:** A fragmentation map of canonical ApoA-I. The peak corresponding to Canonical was isolated and fragmented by ETD, for proteoform characterization and confirmation of ApoA-I identity. The masses of fragment ions observed (fragmentation MS not shown) were then assigned to specific positions where backbone bonds were cleaved. These assignments were then mapped graphically onto the canonical sequence to display fragmentation coverage. A P-score, calculated for this specific map, shows confidence in ApoA-I identification. Moreover, the observed mass of this proteoform corresponded closely to the expected mass for the unmodified canonical ApoA-I backbone. A similar procedure was performed for every proteoform, in order to chemically characterize each species and assign confidence to its characterization. **D-G:** Different proteoforms of ApoA-I observed at the different retention times and SIM scans. Insets show the observed

isotopic distribution of each proteoform in black, compared to the expected distribution, in red.

**Figure S2. Targeted analysis of acylations of ApoA-I.**
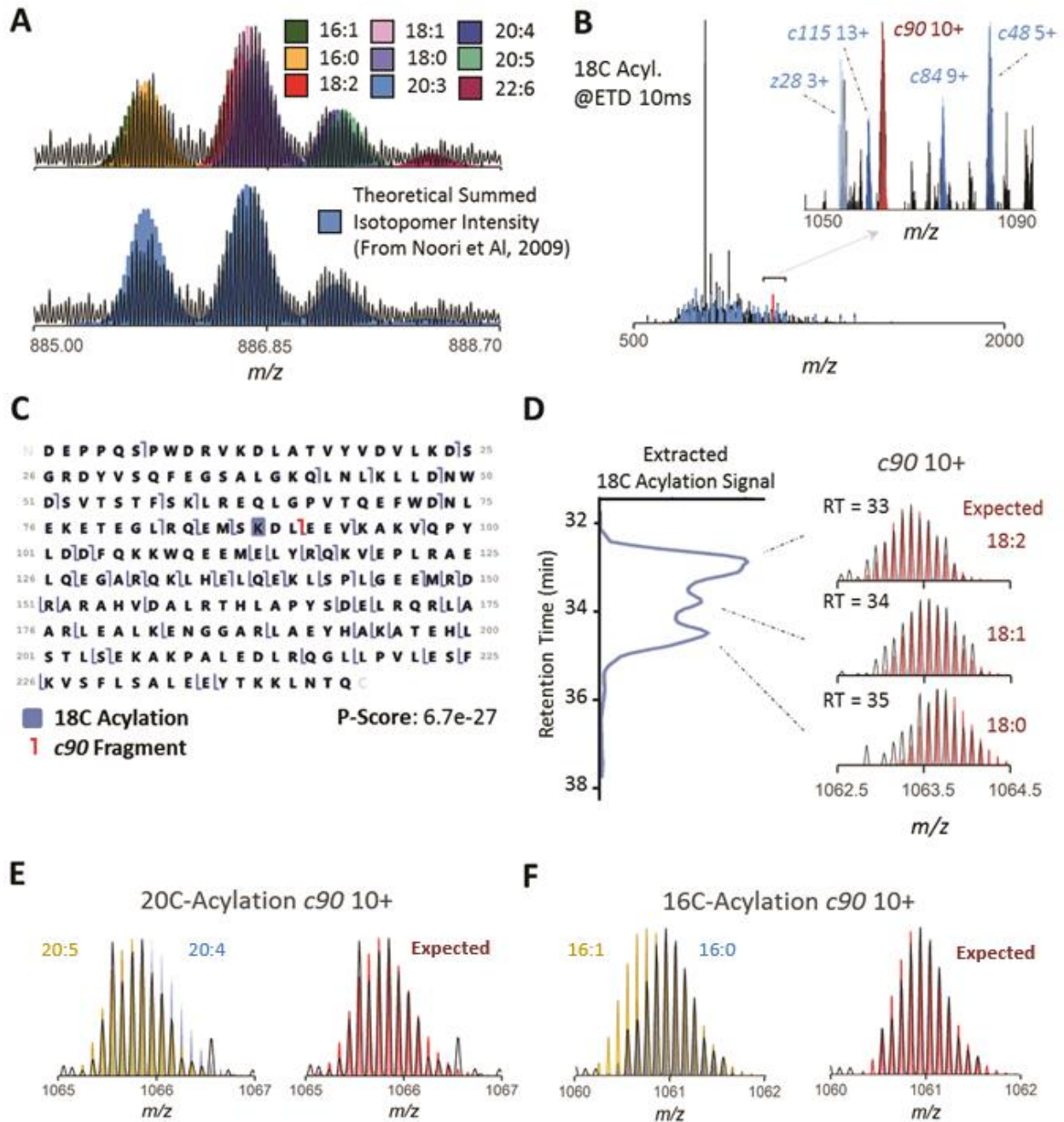
A: mass spectrum of acylated proteoforms on a canonical backbone. In black, an example of an observed spectrum for ApoA-I acylations. On the top panel, different colors denote the isotopic distribution expected for ApoA-I modified by 9 different fatty acids, all

known to be present in HDL. It is noteworthy that, for fatty acids of the same number of carbons, the isotopic distributions of the respective acylated proteoforms largely overlap, making it impossible to separately quantify the proteoforms respective to acids of different unsaturation states (different number of hydrogens) based on intact mass alone. On the bottom panel, the blue trace shows a calculated expected isotopic distribution of these acylations created based on the relative ratios of the different fatty acids in HDL, as described by Noori et al, 2009 [21]. Notably, the relative ratios of these forms are largely consistent with the relative ratios of their respective fatty acids in HDL. **B and C:** typical fragmentation spectrum and map for acylated forms of ApoA-I. For this spectrum, 18-carbon acylations were isolated and submitted to 10ms of ETD. Shown in blue are several common fragments generated by ETD of ApoA-I, and in red, *C90*, generally a high-abundant ETD fragment of ApoA-I and one that contains the acylation site of ApoA-I, K88, highlighted in dark blue on the fragmentation map. This characterization spectrum contained several fragments – in both directions – which included mass-shifts corresponding to a mass of a 18C acylation and confidently characterized this proteoform. **D:** Chromatogram of intact 18C Acylations on a characterization run. When submitted to a shallow gradient, 18C acylations could be separated based on their number of unsaturations. Three chromatographic peaks were observable for the intact form. When analyzing the fragment *C90* across these peaks (inset; black: observed, red: expected), we observed two 2Da shifts (corresponding to one less unsaturation each) overtime, suggesting the chromatographic peaks corresponded to acylations by 18:2, 18:1 and 18:0 fatty acids, respectively. **E and F:** Fragment *C90* analysis of 16C and 20C acylations. Multiple unsaturation forms of both 20C and 16C fatty acids are reported to exist in HDL.

We applied the same shallow gradient and characterization by ETD to these species, however, no chromatographic separation of different unsaturated forms was observed. For 20C acylations, the observed *C90* fragment was a poorer match for both the 20:4 acylation isotopic distribution (in yellow) and the 20:5 one (in blue) than for a combination of both in the relative ratios described in Noori et al, 2009[21] (in red). This suggests that while we could not chromatographically separate the different unsaturated forms of 20C acylations, both 20:4 and 20:5 species are present in the sample. For the 16C acylations both the 16:0 (blue) and the combined isotopic distribution (red) matched better the observed fragment *C90* spectrum than a 16:1 acylation. This is possibly due to the low abundance of 16:1 fatty acids in HDL, at 4% of that of 16:0 fatty acids [21]. It is possible that both species are present in the samples, however, the isotopic distribution is dominated by 16:0 acylations.

Figure S3.Incubation of canonical ApoA-I with palmitate (16:0) and search for acylated proteoforms.
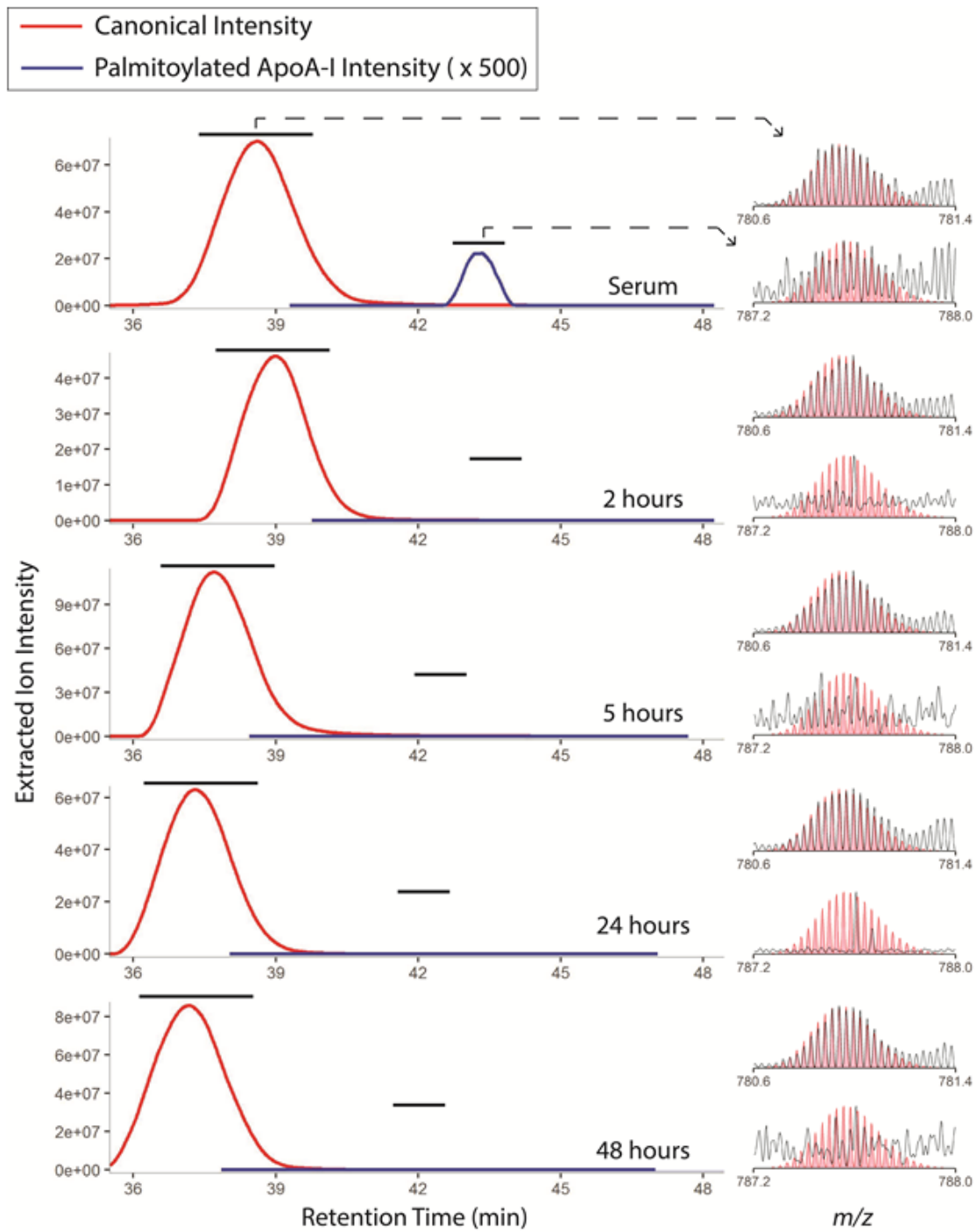
Aliquots from a 2-day incubation of unmodified ApoA-I with palmitic acid were submitted for proteoform search by LCMS. Serum, which contained the palmitoylated proteoform, was run in the same analysis, as a positive control. In red, chromatograms extracted specifically for the canonical proteoform, used as loading control, and in blue chromatograms for palmitoylated ApoA-I. With this search, no significant signal of palmitoylated ApoA-I was observed in incubated fractions. Insets on the right show the spectra observed in each sample at the specific retention times (adjusted for overtime retention-time variation) expected for canonical (top) and palmitoylated ApoA-I (bottom). In red, expected isotopic distributions for these proteoforms are shown in comparison. While a good match with canonical was observed in all samples, only in serum was the isotopic distribution of palmitoylated ApoA-I observed.

Figure S4. Observation and characterization of an allelic form of ApoA-I by LCMS.
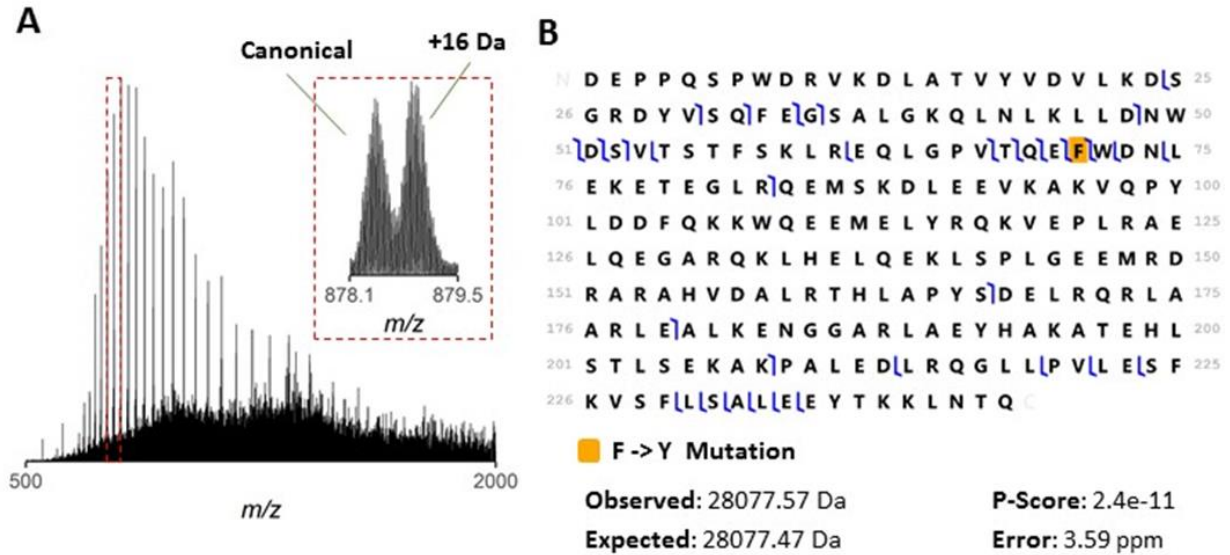
A: a full mass spectrum of ApoA-from a serum sample of the possible heterozygote. Two high-abundant unmodified forms are observed with roughly similar intensities, one 16 Da higher in mass. **B**: HCD-based fragmentation map of the +16 Da form. Fragment ions shifted by 16 Da determine the location of a mas shift at F71. These data are consistent with a F71Y mutation, a common allelic variant of the APOAI gene, with roughly 0.1% prevalence in the population, as described in Haase et al, 2012[22]. Mass error and a P-score are shown for confidence in the characterization.

Figure S5. Mass spectra of the proteoforms of ApoA-II.
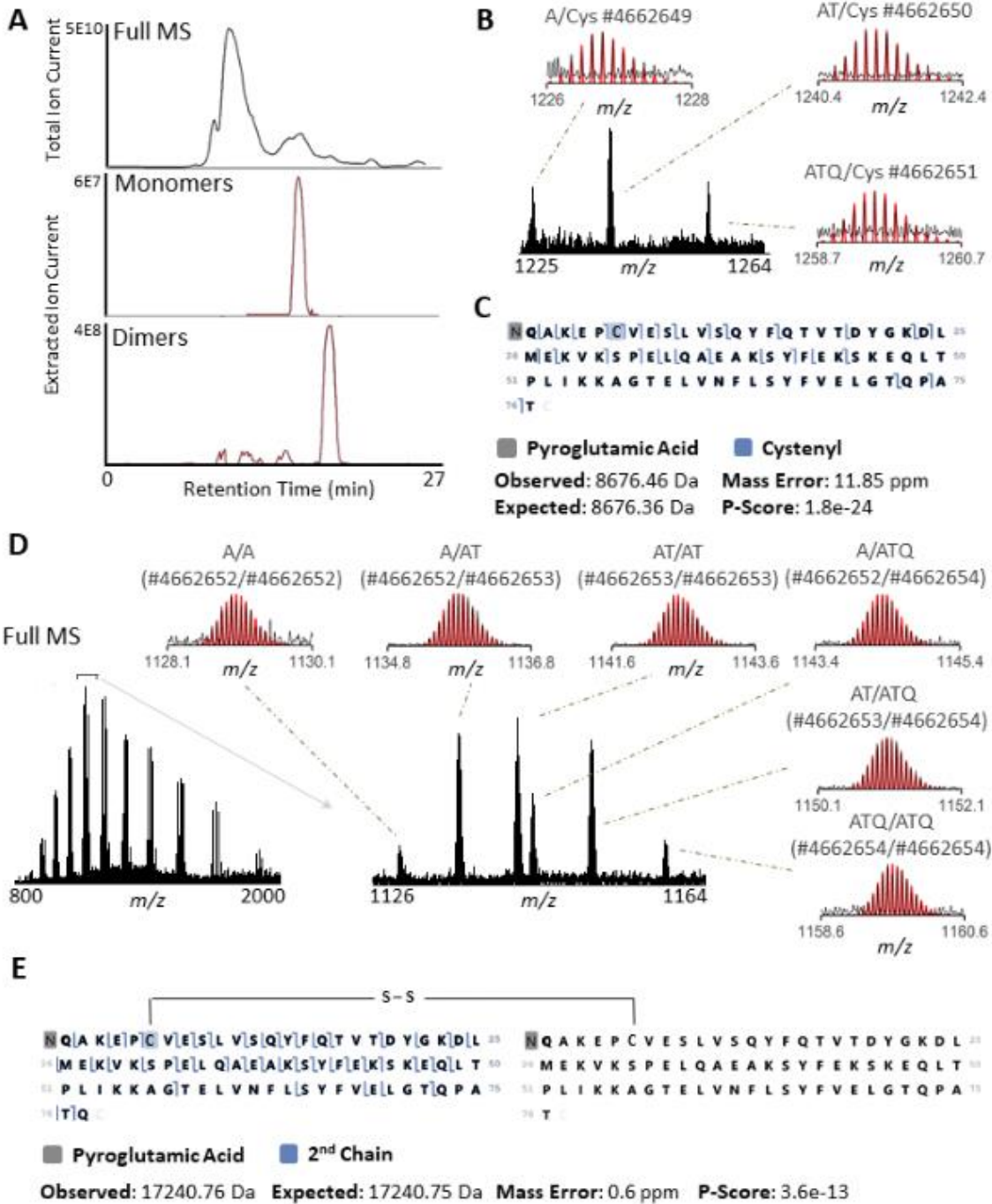
A: Full MS and ApoA-II-specific chromatograms of a typical LCMS run of ApoB-depleted serum. In red, the extracted intensity of all ApoA-II monomers (top) and all dimers (bottom). **B:** Monomers of ApoA-II. Three monomeric forms were observed, containing either a full backbone (ATQ, standing for the last 3 amino-acid residues), one missing the C-terminal glutamine (AT) and one missing two C-terminal residues, glutamine and threonine (A). These backbones were all modified by the mass of a cysteinyl (Cys). Insets show matching of the expected isotopic distributions (red) to the observed data (black). **C:** Fragmentation map of a monomer. The AT/Cys proteoform was isolated and fragmented by ETD. Mapping of fragments to the backbone allowed for the localization of the cystenylation to C6 and of a pyroglutamic acid to the N-terminus (a common modification of N-terminal glutamines). **D:** Dimers of ApoA-II. Left: full MS of all charge states of dimers. Right: zoom on charge state 15. The dimers observed had masses consistent with 2-way combinations of the same 3 backbones as the monomers. Dimers were not cysteinylated, and these proteoforms were labelled based on the truncation stage of their backbones (e.g. A/AT). Insets show matching of the expected isotopic distributions (red) to the observed data (black). **E:** Fragmentation of a dimer. The ATQ/AT proteoform was isolated and submitted to ETD. Fragment mapping showed the mass of a singly truncated backbone (AT) modifying C6 via disulfide bridge.

Figure S6. Parameters of quality control in LCMS-based quantification.
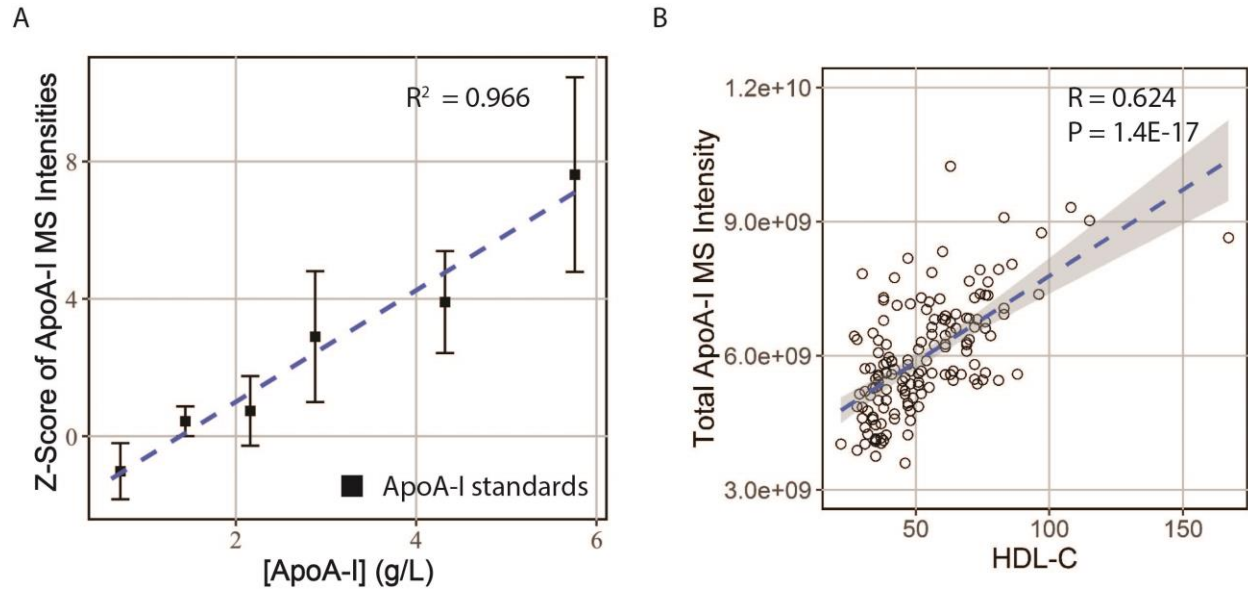
A: Known-concentration commercial ApoA-I standards were injected in random order, at regular intervals within the quantification blocks. The MS intensities of total ApoA-I in these standard samples, standardized within each block are shown in black. Bars show standard deviation. **B:** Individual total ApoA-I MS intensity vs HDL-C. Each point represents one individuals average MS intensity for total ApoA-I and the dashed blue line and shaded area represent a linear regression and confidence interval, respectively. Pearson's r and P-Value are reported.

Figure S7. Co-variance of proteoforms within 150 individuals.

A

C

B

Proteoforms:

	Factor1	Factor2	Factor3
18C Acylations	0.96087	0.05602	-0.0771
16C Acylations	0.95816	0.13264	-0.1188
20C Acylations	0.89789	0.03718	0.21129
DHA Acylations	0.62483	0.02592	-0.0760
20C Acyl. Truncation	0.45128	0.74191	0.14294
16C Acyl. Truncation	0.14324	0.90904	-0.0936
Truncation	-0.3614	0.88467	-0.0567
18C Acyl. Truncation	0.21596	0.88417	-0.0160
DiOxidations	-0.0368	0.02778	0.93877
MonoOxidation	0.01093	-0.056	0.93792
TriOxidation	0.00776	-0.2266	0.63614
ProApoA-I	0.18577	-0.2827	0.43484
Phosphate Adduct	0.31988	-0.1483	-0.0651
Glycation	0.20897	-0.2483	-0.1527
Canonical	0.11428	-0.5474	-0.7774

D

Proteoforms:

	Factor1	Factor2	Factor3
ATQ/ATQ	-0.9908	-0.0203	-0.0506
AT/ATQ	0.1536	-0.9894	-0.1873
AT/AT	1.0322	-0.3285	0.14226
A/ATQ	0.8203	0.22494	0.06434
A/AT	0.7071	0.38421	-0.0932
A/A	0.0717	0.8571	-0.1730
ATQ/Cys	-0.4272	0.12999	0.63813
AT/CYS	0.4984	-0.0194	0.79650

E

Key

-1.0 -0.7 -0.3 0 0.3 0.7 1.0
Pearson's R

A: Scree plot resulting from a factor analysis of ApoA-I proteoform abundance within 150 individuals. Bars: percent variance per factor, points: Eigenvalue. Statistical factors 1, 2 and 3 correspond together to more than 80% of the variation observed, suggesting data on ApoA-I proteoform variation are roughly tri-dimensional. **B:** Loading of each ApoA-I proteoform to each of factors 1, 2 and 3. Highlighted proteoforms are shown circled by the same color in factor vs factor plots on the right. Acylations of ApoA-I, especially the ones carrying a canonical backbone, loaded strongly on factor 1, suggesting that they strongly co-vary within this dataset and that the variation is proportional to statistical factor 1, the one which explains most of the proteoform variation. All forms carrying a truncated backbone loaded positive- and strongly on factor 2 and all oxidized forms loaded positive- and strongly on factor 3. The canonical proteoform was strong- and negatively associated with both factors 2 and 3. These data suggest that statistical factors 1, 2 and 3 may represent the similar mechanistic pathways for the creation of similar proteoforms and that the overall three-dimensionality of the data can be explained by the mechanistic factors behind the co-variation of 3 proteoform families (acylations, truncations and oxidations). Canonical is also associated with two of those dimensions, negatively, suggesting, expectably, that a higher abundance of truncations and oxidations of ApoA-I is associated with a large deficit of the unmodified form. **C:** Scree plot resulting from a factor analysis of ApoA-II proteoform abundance within 150 individuals. Bars: percent variance per factor, points: Eigenvalue. The A/Cys proteoform was removed from this analysis due to the low number of samples in which it was detected. Similar to ApoA-I, the three first factors corresponded together to more than 80% of the variation. **D:** Loading of each ApoA-II proteoform to each of factors 1, 2 and 3. Highlighted proteoforms are

shown circled by the same color in factor vs factor plots on the right. Proteoforms containing an "AT" (singly truncated) backbone all loaded positively on factor 1, with AT/AT loading the strongest, while proteoforms containing only intact backbones (ATQ/ATQ and ATQ/Cys) loaded negatively. On factor 2, proteoforms containing a fully truncated backbone ("A") loaded positively, with A/A loading the strongest, while proteoforms containing an "AT" backbone but not an "A" loaded negatively. On factor 3, only cysteinylated proteoforms (monomers) loaded strongly and positively. These data suggest that most of the variation in ApoA-II can be explained by 3 factors: the first truncation, the second truncation and cysteinylation/dimerization of the backbones. **E:** Heatmap of ApoA-I proteoform variation to ApoA-II proteoform variation correlation factors. Abundances of each proteoform of ApoA-I were compared to abundances of each proteoform of ApoA-II in the 150 individuals studied. A Pearson's r was generated for each correlation observed as well as a correlation P-value. Colors show the strength and sign of each association. Statistical significance, symbolized by an asterisk, was asserted at 5% false discovery rate. Notably, co-varying proteoforms were largely clustered together in this analysis. Moreover, while most proteoforms had no association between the two proteins, truncations of ApoA-I were largely associated with truncations of ApoA-II (both characterized by the loss of a C-terminal glutamine), while canonical ApoA-I was strongly negatively associated with all truncations, as expected from the factor analysis.

Figure S8. Differences in association to phenotype from absolute proteoform intensity to percent proteoform abundance.
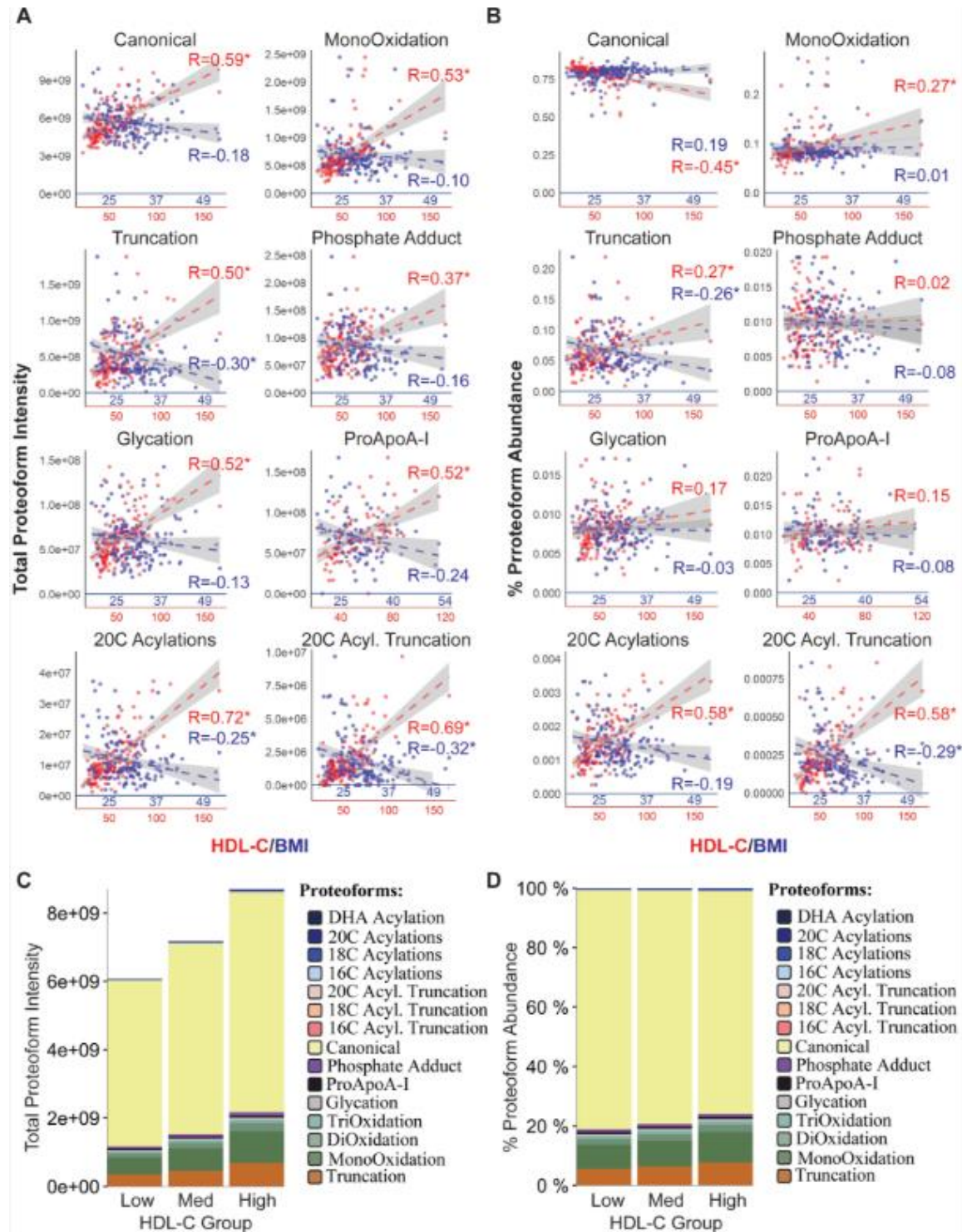
A: Correlative analysis of absolute ApoA-I proteoform intensity to phenotype. In blue, a scatterplot, regression line and Pearson's r of proteoform intensity vs BMI in 150 individuals and in red, the same values, as plotted against HDL-C. Asterisks indicate significance at a 5% false discovery rate. Notably, albeit with different strengths and confidences, all proteoforms of ApoA-I were associated positively with HDL-C and negatively with BMI, expected associations of the total ApoA-I protein concentration to phenotype. **B:** Correlative analysis between the percent contribution of each proteoform to total ApoA-I and phenotype. Percent (or relative) abundances were calculated by dividing each proteoform's intensity to total ApoA-I intensity. Importantly, patterns of association change greatly in this analysis in comparison to absolute intensity, due to the statistical adjusting for the associations of total ApoA-I. The percent abundance of ApoA-I that is unmodified (canonical), for instance, was significantly associated with a lower HDL-C. Moreover, the percent abundance of proteoforms such as the (likely LCMS-generated) phosphate adduct, which is expected not to be endogenous, had no significant association to HDL-C or BMI, suggesting the association observed at absolute intensity was dependent on the association of total proteoform intensity to total ApoA-I. This is not the case for other proteoforms, such as truncated ApoA-I, and acylations of ApoA-I. For these, associations to HDL-C (and in the case of truncations, to BMI as well) were largely independent of the total ApoA-I concentration in individual's sera. **C and D:** Absolute proteoform intensity and percent proteoform abundance, respectively, in the three HDL groups analyzed. Notably, the intensity of all proteoforms is higher in higher HDL-C groups, however, the distribution of proteoform abundances (the proteoform profile) is significantly different across groups: canonical was relatively less prevalent the higher the HDL group and proteoforms such as truncation and acylations were more prevalent in higher HDL individuals.

Figure S9. Correlation coefficient and significance of association of proteoforms percent abundance to binary phenotype.
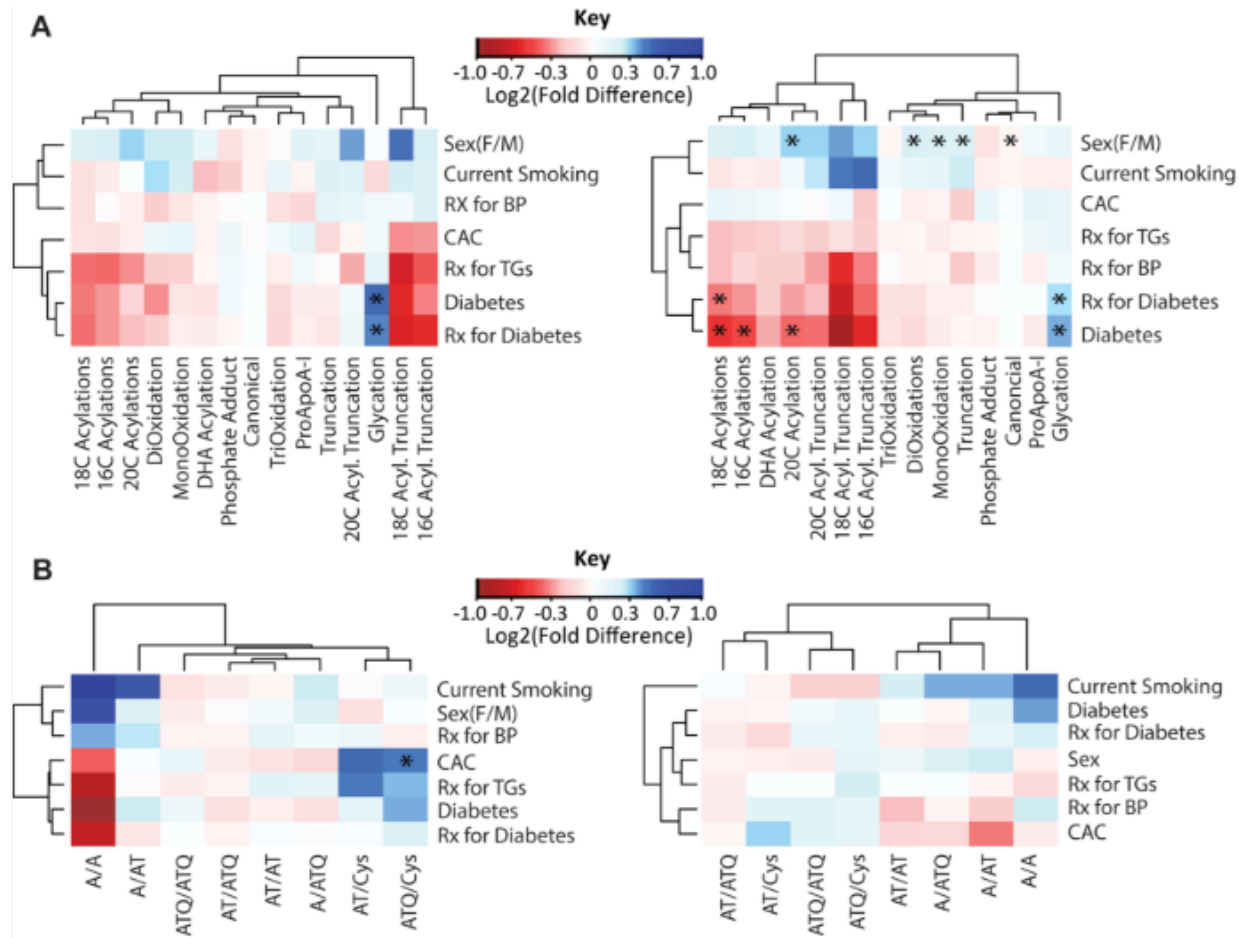
A: ApoA-I proteoform associations. **B:** ApoA-II proteoform associations. Left: CARDIA year 20. Right: CARDIA year 25. Abundances of each proteoform were compared to binary phenotype in the 150 individuals studied. Clustering of both proteoforms and characteristics was unbiased. A P-value was generated from a t-test of the two means. Colors show the Log2 of the fold difference of the means. Statistical significance, symbolized by an asterisk, was asserted at 5% false discovery rate.