Journal of the American Heart Association

ORIGINAL RESEARCH

Derivation of a Protein Risk Score for Cardiovascular Disease Among a Multiracial and Multiethnic HIV+ Cohort

Sandra E. Safo, PhD; Lillian Haine, BA; Jason Baker, MD; Cavan Reilly, PhD; Daniel Duprez, MD, PhD; James D. Neaton, PhD; Mamta K. Jain, MD; Alejandro Arenas-Pinto, PhD; Mark Polizzotto, PhD; Therese Staub, MD; for the ESPRIT, INSIGHT FIRST, SMART and START study groups*

BACKGROUND: Cardiovascular disease risk prediction models underestimate CVD risk in people living with HIV (PLWH). Our goal is to derive a risk score based on protein biomarkers that could be used to predict CVD in PLWH.

METHODS AND RESULTS: In a matched case–control study, we analyzed normalized protein expression data for participants enrolled in 1 of 4 trials conducted by INSIGHT (International Network for Strategic Initiatives in Global HIV Trials). We used dimension reduction, variable selection and resampling methods, and multivariable conditional logistic regression models to determine candidate protein biomarkers and to generate a protein score for predicting CVD in PLWH. We internally validated our findings using bootstrap. A protein score that was derived from 8 proteins (including HGF [hepatocyte growth factor] and interleukin-6) was found to be associated with an increased risk of CVD after adjustment for CVD and HIV factors (odds ratio: 2.17 [95% CI: 1.58–2.99]). The protein score improved CVD prediction when compared with predicting CVD risk using the individual proteins that comprised the protein score. Individuals with a protein score above the median score were 3.10 (95% CI, 1.83–5.41) times more likely to develop CVD than those with a protein score below the median score.

CONCLUSIONS: A panel of blood biomarkers may help identify PLWH at a high risk for developing CVD. If validated, such a score could be used in conjunction with established factors to identify CVD at-risk individuals who might benefit from aggressive risk reduction, ultimately shedding light on CVD pathogenesis in PLWH.

Key Words: cardiovascular disease ■ HIV ■ Olink ■ protein biomarkers ■ proteomics

With effective combination antiretroviral therapy (ART), people living with HIV (PLWH) are living longer, shifting the primary driver of morbidity and mortality for PLWH from opportunistic infections to chronic age-related diseases, such as cardiovascular diseases (CVD).[1-3] PLWH are estimated to have a 1.5- to 2-fold increased risk for CVD when compared with people without HIV.[4] Even in virally suppressed individuals who are HIV-positive, CVD risk is higher than in a population that is HIV negative.[5] Therefore, HIV infection itself is considered a significant risk factor for CVD. Yet CVD screening and prevention methods are currently developed in the general population and do not perform well in a population that is HIV-positive.[6,7] With a high proportion of people with HIV living to older ages, it is imperative to develop CVD risk models tailored to a population that is HIV-positive.

Protein levels could improve prediction of cardiovascular health and as such, the development and use of protein panels could better predict CVD outcomes.[8]

Correspondence to: Sandra E. Safo, PhD, Division of Biostatistics, School of Public Health, University of Minnesota, 420 Delaware St SE, Minneapolis, MN 55455. Email: ssafo@umn.edu

Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.122.027273

*A complete list of the ESPRIT, INSIGHT FIRST, SMART and START study groups can be found in the Appendix in the Supplemental Material.

This article was sent to Pamela N. Peterson, MD, Deputy Editor, for review by expert referees, editorial decision, and final disposition.

For Sources of Funding and Disclosures, see page 9.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- As a strategy to identify potential protein biomarkers that can improve cardiovascular disease risk prediction in people living with HIV, we applied state-of-the-art statistical methods to proteomics data from participants enrolled in 1 of 4 trials by INSIGHT (International Network for Strategic Initiatives in Global HIV Trials).
- We discovered 8 proteins that when combined into a score was strongly associated with cardiovascular disease risk in people living with HIV.

What Are the Clinical Implications?

- A panel of blood biomarkers was found to be more predictive of cardiovascular disease risk than the individual proteins.
- The protein score could be used in conjunction with established factors to identify individuals at risk of cardiovascular disease who might benefit from aggressive risk reduction.

Nonstandard Abbreviations and Acronyms

ENet	elastic net method
CCL25	C-C motif chemokine 25
HGF	hepatocyte growth factor
IDI	integrated discriminant improvement
NRI	net reclassification index
PLA2G7	platelet-activating factor acetylhydrolase
PLSDA	partial least squares discriminant analysis
PLWH	people living with HIV
SCGB3A2	secretoglobin family 3A member
VIP	variable importance in projection

In the general population, a proteomic risk score has been shown to perform better than the Framingham risk score for individuals with stable coronary heart disease.[9,10] Protein biomarkers in inflammation and metabolic regulation pathways such as CRP (C-reactive protein), insulin-like growth factor, and leptin have been found to be related to CVD outcomes.[11] Proteins capture more variability than standard clinical characteristics and thus have the potential to contribute to a more individualized CVD risk assessment.[9] The development of a proteomic risk score for CVD in an ethnically and racially diverse population that is HIV-positive has not yet been explored. We derive a proteomic risk score in PLWH to aid in HIV-specific risk assessment. We hypothesize that by using proteomics data we will identify novel proteins and that a derived protein score will be able to more accurately discriminate CVD cases from controls compared with the individual proteins alone.

METHODS

Requests for data can be made through the INSIGHT (International Network for Strategic Initiatives in Global HIV Trials) website at http://insight-trials.org/. Proposals are revised by the INSIGHT Scientific Steering Committee, all individuals consented to studies of genomics, which covers the institutional review board approval and consent for this project. Codes used in the analysis may be requested from the corresponding author.

Population

Participants were enrolled in 1 of 4 trials conducted by the Community Programs for Clinical Research on AIDS and INSIGHT (International Network for Strategic Initiatives in Global HIV Trials): FIRST (Flexible Initial Retrovirus Suppressive Therapies),[12,13] ESPRIT (Evaluation of Subcutaneous Proleukin [Interleukin-2] in a Randomized International Trial),[14] SMART (Strategies for Management of Antiretroviral Therapy),[15] or START (Strategic Timing of Antiretroviral Therapy).[16] Enrollment in these trials occurred from 1999 to 2013. The results and population under study for each of these trials have been published previously and are reviewed here.[12-16] For all studies, race and ethnicity were self-reported. FIRST enrolled ART-naive participants at sites across the United States who were randomly assigned in a ratio of 1:1:1 to a protease inhibitor strategy, a nonnucleoside reverse transcriptase inhibitor strategy, or a 3-class strategy with both a protease inhibitor and a nonnucleoside reverse transcriptase inhibitor. ESPRIT enrolled patients with HIV who had CD4+ cell counts 300 or more per cubic millimeter and were randomly assigned to receive interleukin-2 plus ART or ART alone from sites across the world including Africa, Asia, Australia, Europe, Israel, North America, South America, and Mexico. SMART randomly assigned people with HIV who had a CD4+ cell count of more than 350 per cubic millimeter to the continuous use of ART or the episodic use of ART from sites across the world including Africa, Asia, Australia, Europe, Israel, North America, South America, and Mexico. START randomly assigned people with HIV who had a CD4+ count of more than 500 cells per cubic millimeter to start ART immediately or to defer it until the CD4+ count decreased to 350 cells per cubic millimeter or until the development of the acquired immunodeficiency syndrome or another condition that dictated the use of

antiretroviral therapy. Individuals were enrolled from Africa, Asia, Australia, Europe, Israel, North America, South America, and Mexico. CVD case was defined as stroke, myocardial infarction, coronary revascularization, coronary artery disease (CAD) requiring surgery, or death from CVD, including unwitnessed deaths with unknown cause.[17,18] For all studies, an end point review committee reviewed documentation provided by clinical sites using prespecified criteria.[18] Myocardial infarction was defined following the universal definition.[19] Strokes were defined using 5 criteria: (1) acute onset with clinically compatible course; (2) computed tomography or magnetic resonance imaging compatible with diagnosis of stroke and current neurologic signs and symptoms; (3) stroke diagnosed as cause of death at autopsy; (4) positive lumbar puncture compatible with subarachnoid hemorrhage; and (5) death certificate or death note from medical record listing stroke as the cause of death. Coronary revascularization or CAD requiring surgery required a medical record from the hospitalization where the procedure was performed. A participant was considered as having a stroke if the first and second criteria were met, the third was met, the first and fourth criteria were met, or the first and fifth criteria were met. Cause of death was coded using documentation of the death provided by the clinical sites using the Coding of Death in HIV system.[20] In all studies, cause of death was coded using documentation of the death provided by the clinical sites using the Coding of Death in HIV system.[20] Each CVD case was matched with 2 controls based on age (±5 years) at baseline, treatment arm of study, and randomization date (±90 days). The matching within treatment arm accounts for ART differences between treatment regimens during the study alongside ART differences between and within the studies included. When controls were not found during the first round of matching, the age and randomization date criteria were widened. All cases and controls consented to the collection of DNA and have genomics data. Stored plasma samples were used to measure protein biomarkers for a case–control CVD substudy.

Proteomics Data From Olink, Normalization, and Quality Control

We used a baseline plasma specimen from consenting individuals to measure protein biomarkers from 5 Olink multiplex panels (Cardiovascular II, Cardiovascular III, Immune Response, Inflammation, and Cardiometabolic). Each panel has 92 target proteins. Detection and sample analysis of proteins from Olink is performed using proximity extension analysis with the target protein detected through high-throughput real-time polymerase chain reaction.[21] The proximity extension analysis assay by Olink has been used in the general population for identifying risk factors for CVD, and inclusion of proteins in the specified panel were chosen based on previously established risk factors and potential risk factors, with collaboration from leading experts in the fields.[22]

The data were intensity normalized, meaning that the data are adjusted to make the median value for each assay on each plate equal to the median for that of the other plates. The data are presented as normalized protein expression values and are reported on the $\log_2$ scale. To monitor the quality of assay performance and the quality of each individual sample, 4 internal controls for each sample are added. There are 2 steps of quality control: (1) each sample plate with above 0.2 normalized protein expression value when evaluated on the SD of the internal controls pass this first step, and (2) samples that deviate <0.3 normalized protein expression value from the median value of the controls for each individual sample pass this step.

Risk Factors

CVD-specific risk factors considered were sex; age; history of CVD at baseline, defined as prior CAD requiring treatment, prior myocardial infarction, prior stroke, or prior CAD requiring surgery; lipid-lowering medication at baseline; blood pressure lowering medication at baseline; self-reported race or ethnicity, which was categorized for SMART, START, and FIRST as either Black, Hispanic, Asian, White, or Other, with ESPRIT as either Black, Asian, White, or Other; and diabetes. Any individual who did not identify as Black was categorized as non-Black. A patient was considered to have diabetes if they received a diagnosis of diabetes requiring drug treatment. One risk factor not included was smoking status at baseline, because it was not measured in ESPRIT. We do consider supplementary analysis on a subgroup with smoking status measurements in Data S1. HIV-specific baseline factors that were considered were ART use, CD4+ values, and HIV RNA values.

Statistical Analysis

Protein Biomarker Identification

We searched for candidate proteins that are able to differentiate CVD cases from controls using bootstrap, partial least squares discriminant analysis (PLSDA), and PLSDA with regularized logistic regression via the elastic[23] net (PLSDA+ ENet). Before bootstrapping, we filtered out proteins with low potential to distinguish between cases and controls via a multivariable conditional logistic regression model predicting CVD case status using each protein while adjusting for age, sex, and Black race and ethnicity. We retained proteins that had P values <0.05. We then created 200 bootstrap

data sets by randomly sampling the data (n=390) with replacement, keeping the ratio of CVD cases and controls. For each bootstrapped set, we fit a 2-component PLSDA model. We used PLSDA variable importance in projection (VIP) scores to identify proteins that contributed most to the differentiation of CVD cases from controls. Following previous studies, we used cutoff values of 1, 1.5, and 2 for the average (over the 200 VIP scores) VIP score. Because there were many proteins with average VIP scores >1, we implemented a logistic regression model with an ENet penalty[24] on each bootstrap data set to reduce the list of potential protein biomarkers. To do this, we ranked proteins based on the frequency of nonzero coefficients out of the 200 ENet models. Proteins ranked in the top 25th percentile were considered as having more potential for differentiating CVD cases from controls. Note that instead of performing all the analyses on the observed data we used bootstrap for stability in the proteins identified. Further, this allowed the data to be used for the protein score and multivariable model developments to be different from the data used for biomarker identification.

Protein Score Development

We developed a protein score to assess the contribution of the identified protein biomarkers in a single combined measure. For each bootstrap data set and each protein, we used a conditional logistic regression model to obtain the log odds ratio of the protein in predicting CVD case status. A weighted mean (over the 200 log-odds ratios from the conditional logistic regression models) was then used to obtain an overall log-odds ratio for each protein. We used the full data set and the weighted mean for each protein to derive the protein score. We estimated the protein score for each individual as the sum of the products of the weighted mean for each protein and the corresponding protein data for that individual.

Model Development, Performance, and Internal Validation

We used multivariable conditional logistic regression models to investigate the following models: (1) baseline, (2) baseline+individual proteins, and (3) baseline+protein score. The baseline model included both HIV-related factors—specifically CD4 and RNA values at baseline—and CVD-related factors—specifically sex, age, body mass index, diabetes status, race, prior history of CVD, and lipid and blood pressure-lowering medication variables all measured at baseline. For prediction estimates, we assessed the area under the curve (AUC) from the receiver operating characteristic curve and reported classification accuracy, sensitivity, and specificity estimates. These estimates were obtained from the optimal cutoff point on the receiver operating characteristic curve (ie, Youden's index). We assessed whether adding the protein score to the baseline model (ie, model 3) improved prediction more than adding each individual protein included in the score to the baseline model (ie, model 2). We assessed the improvements in the performance of model 2 over 1 and of model 3 over 1 using the difference in the area under the receiving operating characteristic curve, the continuous or categorical free net reclassification improvement (NRI),[25,26] and the integrated discriminant improvement (IDI).[25,27] The NRI[25,26] is based on the notion that a valuable new biomarker will tend to increase predicted risks for cases and decrease predicted risks for controls. The IDI[25,27] gives the difference in discrimination slopes between the models compared. For the computations of NRI and IDI, we used the CVD events and predicted probabilities based on the conditional logistic regression models. The continuous NRI was used because that does not depend on prespecified risk categories. In sensitivity analyses, we adjusted for the inflammatory biomarker, high-sensitivity CRP (hsCRP), and the coagulation biomarker, D-dimer, as PLWH with higher levels of these inflammatory biomarkers are more likely to experience fatal CVD events. We carried out separate analyses for individuals who were on ART at baseline, those who did not have any prior history of CVD at baseline, and those with data on smoking status at baseline. In the absence of external validation data, we internally validated our findings using bootstrap. For internal validation, we repeated each step of the model development from candidate protein biomarker selection to protein score development to model development and finally to performance assessment. Please refer to Data S1 for details and Figure S1 for flow chart of our analyses process. We performed all analyses using R software version 3.6.0 (The R Foundation).[28] Statistical significance was ascertained at a 2-sided significance level of 0.05, and because these are exploratory analyses we did not perform any adjustment for multiplicity.

RESULTS

Study Population

Of the 390 individuals with and without CVD included in our analyses, the median (interquartile range) age at baseline was 47 (41–54) years and median (interquartile range) body mass index at baseline was 24.6 (22.2–27.2) kg/m^2. Thirteen percent of participants were female, and 17% were Black. At baseline, 4.9% of participants had a history of a CVD event, 69% were on ART, 18% were taking blood pressure medication, and 19% were taking lipid lowering medication. Baseline demographic and other characteristics of the study participants by case/control status are presented in

Table 1. Baseline Characteristics for Cases and Controls

	Overall	Case	Control	*P* value
N	390	131	259	
Study: N (%)				1.00
Evaluation of Subcutaneous Proleukin [Interleukin-2] in a Randomized International Trial	228 (58.46)	77 (58.78)	151 (58.30)	
Flexible Initial Retrovirus Suppressive Therapies	24 (6.15)	8 (6.11)	16 (6.18)	
Strategies for Management of Antiretroviral Therapy	48 (12.31)	16 (12.21)	32 (12.36)	
Strategic Timing of Antiretroviral Therapy	90 (23.08)	30 (22.90)	60 (23.17)	
Region: N (%)				0.73
Africa	6 (1.54)	2 (1.53)	4 (1.54)	
Asia	12 (3.08)	2 (1.53)	10 (3.86)	
Australia/New Zealand	33 (8.46)	13 (9.92)	20 (7.72)	
Europe	164 (42.05)	55 (41.98)	109 (42.08)	
North/South America	175 (44.87)	59 (45.04)	116 (44.79)	
Male sex: N (%)	341 (87.44)	120 (91.60)	221 (85.33)	0.11
Non-Black race*: N (%)	324 (83.08)	109 (83.21)	215 (83.01)	1.00
History of cardiovascular disease: N (%)	19 (4.87)	16 (12.21)	3 (1.16)	<0.01
Body mass index, kg/m^2: mean (SD)	25.38 (5.03)	25.81 (5.58)	25.16 (4.72)	0.23
Age, y: mean (SD)	47.69 (9.14)	48.60 (9.57)	47.23 (8.89)	0.16
Diabetes diagnosis: N (%)	0.07 (0.25)	0.09 (0.29)	0.06 (0.23)	0.22
CD4+ count, cells/mm^3: mean (SD)	547.79 (215.29)	541.02 (233.76)	551.22 (205.71)	0.66
HIV RNA, copies/mL: mean (SD)	22540.37 (138511.71)	35856.41 (224730.34)	15779.12 (57288.42)	0.18
On antiretroviral therapy: N (%)	271 (69.49)	93 (70.99)	178 (68.73)	0.73
On lipid-lowering treatment: N (%)	72 (19.10)	32 (25.20)	40 (16.00)	0.04
On blood pressure treatment: N (%)	66 (17.51)	33 (25.98)	33 (13.20)	<0.01
Proteomic risk score: mean (SD)	0.00 (1.00)	0.50 (0.95)	−0.25 (0.93)	<0.01
High-sensitivity C-reactive protein, μg/mL: mean (SD)	3.70 (6.23)	4.49 (8.46)	3.30 (4.68)	0.08
D-dimer, μg/mL: mean (SD)	0.41 (0.56)	0.51 (0.87)	0.37 (0.28)	0.02

P values comparing cases and controls for continuous variables are obtained using Kruskal–Wallis tests, and *P* values for binary/categorical variables are obtained using chi-square tests. History of cardiovascular disease at baseline defined as prior coronary artery disease (CAD) requiring treatment, prior myocardial infarction, prior stroke, or prior CAD requiring surgery. On antiretroviral therapy is defined as on therapy at baseline or at study initiation. Race was self-reported race.

*Non-Black race includes people who self-identified as White, Asian, Hispanic, or Other races or ethnicities.

Table 1. Compared with those who did not develop CVD during follow-up, participants who developed CVD during follow-up were more likely at baseline to have a history of CVD at baseline, to be male, and be taking blood pressure- and lipid-lowering medications. Refer to Table S1 for the breakdown of CVD events.

Proteins Identified by PLSDA and PLSDA+ENet

Out of the 459 proteins passing the proteomics quality control step, 107 proteins passed the filtering step with a *P* value <0.05. A protein that appears on multiple panels passed if it met this criterion for at least 1 panel. The number of significant proteins from the individual panels (ignoring overlaps) were as follows: CVD3 (28), immune response (20), cardiometabolic (16), CVD2 (17), and inflammation (26). Table S2 gives all statistically significant proteins used in PLSDA.

Twenty-nine proteins were identified as being more able to distinguish between CVD cases and controls with average VIP scores >1 (Table S2). A protein that appeared on multiple panels is selected if it has VIP>1 for at least 1 panel. After using a logistic regression with an ENet penalty on the 29 proteins identified with average VIP scores >1, 8 proteins (FAM3B, ITGA11 [integrin α11], IL6 [interleukin-6], HGF [hepatocyte growth factor], CCL25 [C-C motif chemokine 25], gastrotropin, PLA2G7 [platelet-activating factor acetylhydrolase], SCGB3A2 [secretoglobin family 3A member]) were identified as

having more potential for differentiating CVD cases from controls. Table S3 gives the count and proportions of times these 8 proteins had nonzero coefficients from the ENet model across all 200 bootstrap sets. A protein score consisting of these 8 proteins was then developed. When the average VIP score cutoff criterion was set to 1.5 or 2, only 1 protein (gastrotropin) was selected. Figure S2 gives the correlation of these protein biomarkers. The correlation between the proteins ranged from −0.13 to 0.36, and none of the proteins was highly correlated (ie, >0.5) with the other proteins.

Comparison of Baseline and Protein Models

A multivariable conditional logistic regression model predicting CVD case status using the protein score with adjustment for CVD- and HIV-related risk factors was statistically significant (P<0.05 [95% CI, 1.58–2.99]; Figure and Table S4). A 1-SD increase in the protein score was associated with an odds ratio (OR) for CVD of 2.17. The proteins FAM3B, IL6, HGF, gastrotropin, PLA2G7, and SCGB3A2 were each statistically significant (P<0.05) when added to the baseline model (Table 2) and their corresponding ORs were lower than the OR for the protein score in the protein score+baseline model (we are able to directly compare the ORs of the individual proteins to that of the protein score because these proteins and the protein score have each been standardized to have variance 1). Further, IL6 and PLA2G7 had AUC values that were somewhat higher than the other proteins (Table 3), and similar to the AUC from the protein score+baseline model. However, the NRIs for IL6 and PLA2G7 were 0.40 (95% CI, 0.19–0.61) and 0.28 (95% CI, 0.06–0.49), respectively, as compared with 0.66 (95% CI, 0.46–0.86) for the baseline+protein score model. The protein score+baseline model had an AUC of 0.73 (95% CI, 0.65–0.81), which was a 5.8% improvement in AUC over the baseline model (AUC=0.69 [95% CI, 0.60–0.78]), or a 17.4% (0.73–0.69)/(0.73–0.50) improvement in prediction ability (Table 3). The prediction accuracy, sensitivity, and specificity estimates (evaluated at the optimal cutoff point) for the protein score+baseline model were 0.75, 0.78, and 0.73, respectively, compared with 0.65, 0.81, and 0.57 for the baseline model. The NRI (Table 4) was 0.66 (95% CI, 0.46–0.86); this was driven more by net proportion of cases assigned a higher risk (NRI of cases is 0.41 [95% CI, 0.25–0.57]) than in net proportion of controls assigned a lower risk (NRI of controls is 0.25 [95% CI, 0.13–0.37]). The IDI was estimated to be statistically significant (IDI = 0.09 [95% CI, 0.06–0.14]). Three of the 8

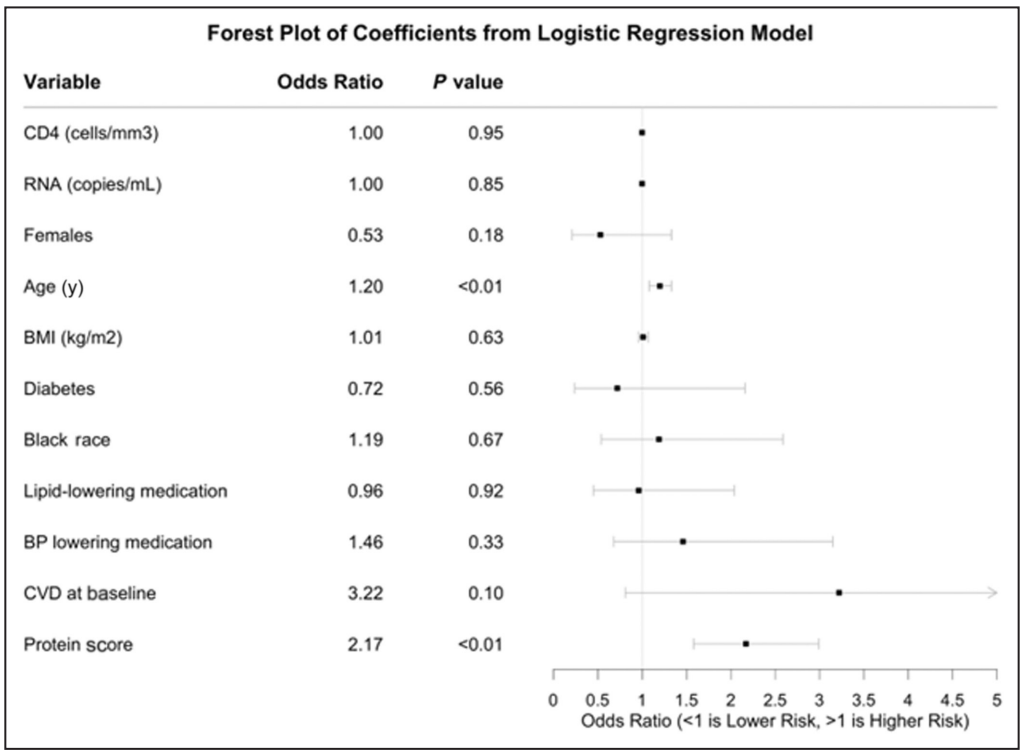


Figure. Conditional logistic regression model of CVD on standardized protein score.
Odds ratios for women compare women to men; Black race is compared with non-Black race or ethnicity. BMI indicates body mass index; BP, blood pressure; CD4, CD4+ count at baseline; and CVD, cardiovascular disease. CVD at baseline was defined as prior coronary artery disease (CAD) requiring treatment, prior myocardial infarction, prior stroke, or prior CAD requiring surgery. Race was self-reported race.

Table 2. Conditional Logistic Regression Model of CVD on Each of the Standardized Protein Used to Develop the Protein Score

Variable	Odds ratio	SE	*P*-value	Lower CI	Upper CI
Baseline					
+ FAM3B	1.48	0.14	0.01	1.12	1.95
+ Integrin α11	0.85	0.13	0.20	0.66	1.09
+ Interleukin-6	1.54	0.13	0.00	1.18	2.00
+ Hepatocyte growth factor	1.63	0.15	0.00	1.22	2.17
+ C-C motif chemokine 25	1.26	0.12	0.06	0.99	1.61
+ Gastrotropin	0.72	0.14	0.02	0.55	0.95
+ Platelet-activating factor acetylhydrolase	1.48	0.14	0.01	1.12	1.95
+ Secretoglobin family 3A member	1.32	0.13	0.03	1.03	1.69

Here, n=375 (cases=126 and controls=249) as there were 15 samples with missing values. Baseline is a model with the following variables: CD4, RNA, sex, age, body mass index, diabetes status at baseline, prior history of CVD, lipid-lowering medication, and blood pressure-lowering medication. CVD indicates cardiovascular disease.

proteins in the protein score (FAM3B, ITGA11, IL6) were found on the immune response Olink panel; 2 (HGF and CCL25) were from the inflammation panel; gastrotropin, SCGB3A2, and PLA2G7 were from the CVD2, CVD3, and cardiometabolic panels, respectively. A description of each protein is found in Data S1.

Categorization of Protein Score and Association With CVD Risk

To further assess the impact of the protein score on CVD risk, we categorized the protein score and considered associations of the categorized score with CVD. When the score was categorized above and below the median protein score, individuals with a protein score above the median were 3.1 times more likely to have CVD compared with individuals with a score below the median (Table S5). When the score was dichotomized as those individuals with scores in the top 25% versus those with scores in the bottom 75%, we found that those with scores in the top 25% were 2.9 times more likely to have CVD compared with those with scores in the bottom 75% (Table S5).

DISCUSSION

Our assessment of the contribution of protein biomarkers to CVD risk prediction showed that a protein risk score developed using 8 proteins—FAM3B, ITGA11, IL6, HGF, CCL25, gastrotropin, PLA2G7,

Table 3. Incremental Contribution of Individual Proteins and Protein Score to CVD Risk When Added to Baseline Model (n=375)

Model	AUC	Lower CI	Upper CI	Change in AUC from baseline	NRI among cases	NRI among controls	Overall NRI
Baseline*	0.69	0.60	0.78	…	…	…	…
+ FAM3B	0.71	0.63	0.80	0.02	0.22 (0.05, 0.39)	0.04 (−0.08, 0.17)	0.27 (0.06, 0.48)
+ Integrin α11	0.69	0.61	0.78	0.004	0.13 (−0.05, 0.30)	−0.00 (−0.13, 0.12)	0.12 (−0.09, 0.34)
+ Interleukin-6	0.73	0.65	0.81	0.04	0.24 (0.07, 0.41)	0.17 (0.04, 0.29)	0.40 (0.19, 0.61)
+ Hepatocyte growth factor	0.71	0.62	0.80	0.02	0.18 (0.00, 0.35)	0.12 (−0.01, 0.24)	0.29 (0.08, 0.50)
+ C-C motif chemokine 25	0.71	0.62	0.80	0.02	0.06 (−0.11, 0.24)	0.14 (0.02, 0.26)	0.20 (−0.01, 0.42)
+ Gastrotropin	0.70	0.62	0.79	0.01	0.22 (0.05, 0.39)	0.12 (−0.01, 0.24)	0.34 (0.13, 0.55)
+ Platelet-activating factor acetylhydrolase	0.72	0.64	0.80	0.03	0.14 (−0.03, 0.32)	0.13 (0.01, 0.26)	0.28 (0.06, 0.49)
+ Secretoglobin family 3A member	0.70	0.62	0.79	0.01	0.18 (0.00, 0.35)	0.16 (0.03, 0.28)	0.33 (0.12, 0.54)
+ Protein score from all 8 proteins	0.73	0.65	0.81	0.04	0.41 (0.25, 0.57)	0.25 (0.13, 0.37)	0.66 (0.46, 0.86)

Baseline* is a model with the following variables: CD4, RNA, sex, age, body mass index, diabetes status at baseline, prior history of cardiovascular disease, lipid-lowering medication, and blood pressure-lowering medication. The proteins FAM3B, IL6, HGF, gastrotropin, PLA2G7, and SCGB3A2 were each statistically significant ($P<0.05$) when added to the baseline model. Protein score was statistically significant ($P<0.01$) when added to the baseline model. Odds ratio for protein score: 2.17 (CI, 1.58–2.97). AUC indicates area under the curve; CCL25, C-C motif chemokine 25; FAM3B, protein FAM3B; HGF, hepatocyte growth factor; IL6, interleukin-6; NRI, net reclassification index; PLA2G7, platelet-activating factor acetylhydrolase; and SCGB3A2, secretoglobin family 3A member.

Table 4. Reclassification Table Using Category-Free NRI. Model With Protein Score in Addition to Baseline Variables

	All	Reclassified upwards n (proportion)	Reclassified downwards N (proportion)	NRI	
Number of cases	126	88.96 (0.71)	37.04 (0.29)	Among cases	0.41 (0.25–0.57)
Number of controls	249	92.88 (0.37)	153.88 (0.62)	Among controls	0.25 (0.13–0.37)
Overall NRI (CI)					0.66 (0.46–0.86)

Net reclassification index (NRI) of cases is greater than NRI of controls. This suggests that addition of the protein score helps increase predicted risk for those with cases more than it decreases predicted risk for controls.

SCGB3A2—was associated with CVD risk. A description of current literature for each protein is found in Data S1. Of these 8 proteins, IL6 is the only one to have previously been associated with CVD in both PLWH and the general population.[17,29,30] HGF and PLA2G7 have both been found to be associated with coronary heart disease in the general population.[31-34] The results here indicate that both HGF and PLA2G7 are potentially also associated with CVD in PLWH. In the general population, dysregulation of FAM3B is associated with diabetes, a known risk factor for CVD.[35] CCL25 is related to T cells, which are involved in the development and progression of CVDs in the general population.[36] One novel protein is ITGA11, which has not been extensively studied in humans.[37,38] Another novel protein is gastrotropin, a member of the fatty acid-binding protein family, which is thought to serve an integral role in metabolic function.[39,40] SCGB3A2 has been found in a case–control study in a Korean population to contribute to susceptibility to asthma.[41,42]

The protein score predicted CVD better when added to both HIV and CVD risk factors compared with models with the individual proteins and HIV and CVD risk factors. The model with the protein score and HIV and CVD risk factors showed better prediction performance compared with the HIV and CVD factors only model. When the protein risk score was dichotomized, our analyses showed that individuals with a score above the median score were 3 times more likely to develop CVD, which is illustrated in Table S5. The inclusion of D-dimer and hsCRP in the baseline model did not alter our findings, illustrated in Tables S6 and S7, respectively. The following subgroup analyses were done: the first, restricted to individuals on ART at baseline, is shown in Table S8, the second, restricted to individuals with no prior history of CVD, is shown in Table S9, and the third, restricted to individuals with smoking data at baseline, is shown in Tables S10 and S11, all again demonstrated the ability of the protein score to discriminate between CVD cases and controls. Model performance statistics from the sensitivity analyses were comparable to model performance statistics from the full population. Using resampling techniques, we were able to internally validate our findings; internal validation results are shown in Table S12. Please refer to Data S1 for results on sensitivity, subgroup, and internal validation analyses.

A proteomic risk score has been shown to perform better than the Framingham risk score for individuals with stable coronary heart disease.[9,10] Proteins are able to capture variability beyond standard clinical characteristics and thus could contribute to an individualized CVD risk assessment.[9] This article identified candidate proteins and derived a proteomic risk score in PLWH to aid in HIV-specific risk assessment.

The incremental AUC (5.8%), the NRI (0.66), and the IDI (0.09) of the protein score, based on published risk estimates, indicate a moderate improvement in risk prediction. Similar improvements in the AUC were shown in a 9-protein risk score developed and validated for coronary heart disease in Ganz et al.[10] with similar improvement in risk prediction with a C statistic (or AUC) that ranges from 0.64 for refit Framingham to 0.71 for refit Framingham plus 9-protein risk score, which is a 10.9% improvement in risk prediction. This is similar to our improvement from a 0.69 for the baseline model to 0.73 for baseline plus our proteomic risk score, which is a 5.8% improvement in risk prediction.

Our article has several strengths. One is the use of an ethnically and racially diverse population, which provides information on understudied groups. However, there was not sufficient sample size to evaluate if the predictive value of the protein score was similar or different by race and ethnicity. A second strength is the use of a population that is HIV-positive. PLWH are at an increased risk for CVDs with substantial variability in CVD risk that is left unexplained by established risk factors. This study begins to characterize some of the heterogeneity within PLWH. Third is the use of a protein score because of its potential clinical relevance. Polygenic risk scores are often criticized for lack of clinical relevance due to the need to use complex machinery and statistical methods to genotype every individual, which is not yet readily available in the clinical setting. Comparatively the measurement of proteins and this protein score, which involves only 8 proteins, is potentially more clinically applicable as it is more easily measurable and thus more clinically relevant. A final strength is the use of state-of-the-art statistical methods for biomarker identification.

One limitation is a small (n=390) study sample size, limiting power for detecting modest associations and inhibiting our ability to validate the risk score on an independent sample, instead we rely on resampling techniques for validation. We did not have controls who were HIV negative, so it is unclear whether the increased risk is specific to HIV infection. Further, only 69% of participants were on ART so it is likely that our findings will change if all participants were on ART. We note that because we matched cases and controls within treatment arms, ART differences between treatment regimens during the study are minimized. When we restricted our analysis to participants on ART at baseline (n=269), the protein score was again statistically significant, and improved prediction beyond the baseline model. We observed that unlike in the combined data where IL6 was likely driving the effect of the protein score, in this subpopulation, no individual protein had an AUC that was comparable to the protein score+baseline model. As another limitation, we did not account for social determinants of health conditions such as socioeconomic status and neighborhood environment variables, which have been suggested to affect CVD in the general population.[43] Finally, current smoking status at baseline was not measured in all the studies. As such, we did supplementary subgroup analyses on the individuals with current smoking status measurements. However, these results are limited due to small sample size.

CONCLUSIONS

We find that a proteomic risk score developed in a multiethnic and multiracial cohort that is HIV-positive is a potentially beneficial approach for capturing heterogeneity above and beyond established risk factors. We developed a protein risk score that has high potential for benefit in CVD risk prediction in PLWH. We also provide a statistical approach to protein score development that can be applied for more individualized risk prediction.

ARTICLE INFORMATION

Received June 22, 2022; accepted February 28, 2023.

Affiliations

University of Minnesota, Minneapolis, MN (S.E.S., L.H., C.R., D.D., J.D.N.); Hennepin County Medical Center, Minneapolis, MN (J.B.); University of Texas Southwestern, Dallas, TX (M.K.J.); MRC Clinical Trials Unit at University College London Institute of Clinical Trials & Methodology, London, UK (A.A.-P.); Australian National University, Canberra, Australia (M.P.); and Luxembourg Hospital Center, Luxembourg City, Luxembourg (T.S.).

Sources of Funding

This study was funded by National Institutes of Health (NIH) grants UM1-AI068641, UM1-AI120197, and U01-AI136780 (for START); NIH grants UM1-AI068641 and U01-AI46957 (for ESPRIT); NIH grants U01-AI042170 and U01AI046362 (for FIRST); and NIH grants UM1-AI068641, U01-AI046362, and U01-AI042170 (for SMART). Sandra Safo was partially supported by grant numbers 5KL2TR002492-04 from NIH and 13XS134T11 from Leidos and NIH. Lillian Haine was supported by the NIH under award number T32HL129956 and in part by the NIH's National Center for Advancing Translational Sciences (NCATS) grants TL1R002493 and UL1TR002494. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH's NCATS.

Disclosures

None.

Supplemental Material

Data S1–S2
Tables S1–S12
Figures S1–S2

REFERENCES

1. Smith CJ, Ryom L, Weber R, Morlat P, Pradier C, Reiss P, Kowalska JD, de Wit S, Law M, el Sadr W, et al; D:A:D Study Group. Trends in underlying causes of death in people with HIV from 1999 to 2011 (D:a:D): a multicohort collaboration. *Lancet*. 2014;384:241–248. doi: 10.1016/S0140-673660604-8
2. Farahani M, Mulinder H, Farahani A, Marlink R. Prevalence and distribution of non-AIDS causes of death among HIV-infected individuals receiving antiretroviral therapy: a systematic review and meta-analysis. *Int J STD AIDS*. 2017;28:636–650. doi: 10.1177/0956462416632428
3. Drozd DR, Kitahata MM, Althoff KN, Zhang J, Gange SJ, Napravnik S, Burkholder GA, Mathews WC, Silverberg MJ, Sterling TR, et al. Increased risk of myocardial infarction in HIV-infected individuals in North America compared with the general population. *J Acquir Immune Defic Syndr*. 2017;75:568–576. doi: 10.1097/QAI.0000000000001450
4. Shah ASV, Stelzle D, Lee KK, Beck EJ, Alam S, Clifford S, Longenecker CT, Strachan F, Bagchi S, Whiteley W, et al. Global burden of atherosclerotic cardiovascular disease in people living with HIV systematic review and meta-analysis. *Circulation*. 2018;138:1100–1112. doi: 10.1161/CIRCULATIONAHA.117.033369
5. Longenecker CT, Sullivan C, Baker JV. Immune activation and cardiovascular disease in chronic HIV infection. *Curr Opin HIV AIDS*. 2016;11:216–225. doi: 10.1097/COH.0000000000000227
6. Thompson-Paul AM, Lichtenstein KA, Armon C, Palella FJ Jr, Skarbinski J, Chmiel JS, Hart R, Wei SC, Loustalot F, Brooks JT, et al. Cardiovascular disease risk prediction in the HIV outpatient study. *Clin Infect Dis*. 2016;63:1508–1516. doi: 10.1093/cid/ciw615
7. Triant VA, Perez J, Regan S, Massaro JM, Meigs JB, Grinspoon SK, D'Agostino RB Sr. Cardiovascular risk prediction functions underestimate risk in HIV infection. *Circulation*. 2018;137:2203–2214. doi: 10.1161/CIRCULATIONAHA.117.028975
8. Lindsey ML, Mayr M, Gomes AV, Delles C, Arrell DK, Murphy AM, Lange RA, Costello CE, Jin YF, Laskowitz DT, et al. Transformative impact of proteomics on cardiovascular health and disease: a scientific statement from the American Heart Association. *Circulation*. 2015;132:852–872. doi: 10.1161/CIR.0000000000000226
9. Sze SK. Artificially intelligent proteomics improves cardiovascular risk assessment. *EBioMedicine*. 2019;40:23–24. doi: 10.1016/j.ebiom.2019.01.014
10. Ganz P, Heidecker B, Hveem K, Jonasson C, Kato S, Segal MR, Sterling DG, Williams SA. Development and validation of a protein-based risk score for cardiovascular outcomes among patients with stable coronary heart disease. *JAMA*. 2016;315:2532–2541. doi: 10.1001/jama.2016.5951
11. Ho JE, Lyass A, Courchesne P, Chen G, Liu C, Yin X, Hwang SJ, Massaro JM, Larson MG, Levy D. Protein biomarkers of cardiovascular disease and mortality in the community. *J Am Heart Assoc*. 2018;7:7. doi: 10.1161/JAHA.117.008108
12. MacArthur RD, Chen L, Mayers DL, Besch CL, Novak R, van den Berg-Wolf M, Yurik T, Peng G, Schmetter B, Brizz B, et al. The rationale and design of the CPCRA (Terry Beirn Community Programs for Clinical Research on AIDS) 058 FIRST (Flexible Initial Retrovirus Suppressive Therapies) trial. *Control Clin Trials*. 2001;22:176–190. doi: 10.1016/S0197-2456(01)00111-8
13. MacArthur RD, Novak RM, Peng G, Chen L, Xiang Y, Hullsiek KH, Kozal MJ, van den Berg-Wolf M, Henely C, Schmetter B, et al. A comparison of three highly active antiretroviral treatment strategies

consisting of non-nucleoside reverse transcriptase inhibitors, protease inhibitors, or both in the presence of nucleoside reverse transcriptase inhibitors as initial therapy (CPCRA 058 FIRST study): a long-term randomised trial. *Lancet*. 2006;368:2125–2135. doi: 10.1016/S0140-6736(06)69861-9
14. Emery S, Abrams DI, Cooper DA, Darbyshire JH, Lane HC, Lundgren JD, Neaton JD, ESPRIT Study Group. The Evaluation of Subcutaneous Proleukin® (interleukin-2) in a Randomized International Trial: rationale, design, and methods of ESPRIT. *Control Clin Trials*. 2002;23:198–220. doi: 10.1016/S0197-2456(01)00179-9
15. Strategies for Management of Antiretroviral Therapy (SMART) Study Group, El-Sadr WM, Lundgren J, Neaton JD, Gordin F, Abrams D, Arduino RC, Babiker A, Burman W, Clumeck N, et al. CD4+ count–guided interruption of antiretroviral treatment. *N Engl J Med*. 2006;355:2283–2296. doi: 10.1056/nejmoa062360
16. INSIGHT START Study Group, Lundgren JD, Babiker AG, Gordin F, Emery S, Grund B, Sharma S, Avihingsanon A, Cooper DA, Fätkenheuer G, et al. Initiation of antiretroviral therapy in early asymptomatic HIV infection. *N Engl J Med*. 2015;373:795–807. doi: 10.1056/NEJMoa1506816
17. Nordell AD, McKenna M, Borges AH, Duprez D, Neuhaus J, Neaton JD. Severity of cardiovascular disease outcomes among patients with HIV is related to markers of inflammation and coagulation. *J Am Heart Assoc*. 2014;3. doi: 10.1161/JAHA.114.000844
18. Lifson AR, INSIGHT Endpoint Review Committee Writing Group, Belloso WH, Davey RT, Duprez D, Gatell JM, Hoy JF, Krum EA, Nelson R, Pedersen C, et al. Development of diagnostic criteria for serious non-AIDS events in HIV clinical trials. *HIV Clin Trials*. 2010;11:205–219. doi: 10.1310/HCT1104-205
19. Thygesen K, Alpert JS, Jaffe AS, Chaitman BR, Bax JJ, Morrow DA. White HD; Executive Group on behalf of the Joint European Society of Cardiology (ESC)/American College of Cardiology (ACC)/American Heart Association (AHA)/World Heart Federation (WHF) Task Force for the Universal Definition of Myocardial Infarction. Fourth universal definition of myocardial infarction (2018). *Circulation*. 2018;138:e618–e651. doi: 10.1161/CIR.0000000000000617
20. Kowalska JD, Friis-Møller N, Kirk O, Bannister W, Mocroft A, Sabin C, Reiss P, Gill J, Lewden C, Phillips A, et al. The Coding Causes of Death in HIV (CoDe) Project: initial results and evaluation of methodology. *Epidemiology*. 2011;22:516–523. doi: 10.1097/EDE.0B013E31821B5332
21. Assarsson E, Lundberg M, Holmquist G, Björkesten J, Thorsen SB, Ekman D, Eriksson A, Rennel Dickens E, Ohlsson S, Edfeldt G, et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One*. 2014;9:e95192. doi: 10.1371/journal.pone.0095192
22. Olink CVD III. protein biomarker panel for cardiovascular disease studies. Accessed April 26, 2021 https://www.olink.com/products/target/cvd-iii-panel/.
23. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33:1–22. doi: 10.18637/jss.v033.i01
24. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J Roy Stat Soc, Series B (Statistical Methodology)*. 2005;67:301–320. doi: 10.1111/j.1467-9868.2005.00503.x. http://www.jstor.org/stable/3647580
25. Pencina MJ, d'Agostino RB, d'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008;27:157–172. doi: 10.1002/sim.2929
26. Pencina MJ, D'Agostino RB, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med*. 2011;30:11–21. doi: 10.1002/sim.4085
27. Kerr KF, McClelland RL, Brown ER, Lumley T. Evaluating the incremental value of new biomarkers with integrated discrimination improvement. *Am J Epidemiol*. 2011;174:364–374. doi: 10.1093/aje/kwr086
28. R Core Team. R: a language and environment for statistical computing. 2019.
29. Duprez DA, Neuhaus J, Kuller LH, Tracy R, Belloso W, De Wit S, Drummond F, Lane HC, Ledergerber B, Lundgren J, et al. Inflammation, coagulation and cardiovascular disease in HIV-infected individuals. *PLoS ONE*. 2012;7:e44454. doi: 10.1371/journal.pone.0044454
30. Hsu DC, Ma YF, Hur S, Li D, Rupert A, Scherzer R, Kalapus SC, Deeks S, Sereti I, Hsue PY. Plasma IL-6 levels are independently associated with atherosclerosis and mortality in HIV-infected individuals on suppressive antiretroviral therapy. *AIDS*. 2016;30:2065–2074. doi: 10.1097/QAD.0000000000001149
31. Bielinski SJ, Berardi C, Decker PA, Larson NB, Bell EJ, Pankow JS, Sale MM, Tang W, Hanson NQ, Wassel CL, et al. Hepatocyte growth factor demonstrates racial heterogeneity as a biomarker for coronary heart disease. *Heart*. 2017;103:1185–1193. doi: 10.1136/heartjnl-2016-310450
32. Bell EJ, Decker PA, Tsai MY, Pankow JS, Hanson NQ, Wassel CL, Larson NB, Cohoon KP, Budoff MJ, Polak JF, et al. Hepatocyte growth factor is associated with progression of atherosclerosis: the Multi-Ethnic Study of Atherosclerosis (MESA). *Atherosclerosis*. 2018;272:162–167. doi: 10.1016/j.atherosclerosis.2018.03.040
33. Hurt-Camejo E, Camejo G, Peilot H, Öörni K, Kovanen P. Phospholipase A2 in vascular disease. *Circ Res*. 2001;89:298–304. doi: 10.1161/hh1601.095598
34. Caslake MJ, Packard CJ. Lipoprotein-associated phospholipase A2 (platelet-activating factor acetylhydrolase) and cardiovascular disease. *Curr Opin Lipidol*. 2003;14:347–352. doi: 10.1097/00041433-200308000-00002
35. Zhang W, Chen S, Zhang Z, Wang C, Liu C. FAM3B mediates high glucose-induced vascular smooth muscle cell proliferation and migration via inhibition of miR-322-5p. *Sci Rep*. 2017;7:7. doi: 10.1038/s41598-017-02683-3
36. Jung MK, Shin EC. Aged T cells and cardiovascular disease. *Cell Mol Immunol*. 2017;14:1009–1010. doi: 10.1038/cmi.2017.111
37. Talior-Volodarsky I, Connelly KA, Arora PD, Gullberg D, McCulloch CA. α11 integrin stimulates myofibroblast differentiation in diabetic cardiomyopathy. *Cardiovasc Res*. 2012;96:265–275. doi: 10.1093/cvr/cvs259
38. Romaine A, Sørensen IW, Zeltz C, Lu N, Erusappan PM, Melleby AO, Zhang L, Bendiksen B, Robinson EL, Aronsen JM, et al. Overexpression of integrin α11 induces cardiac fibrosis in mice. *Acta Physiologica*. 2018;222:222. doi: 10.1111/apha.12932
39. Thumser AE, Moore JB, Plant NJ. Fatty acid binding proteins: tissue-specific functions in health and disease. *Curr Opin Clin Nutr Metab Care*. 2014;17:124–129. doi: 10.1097/MCO.0000000000000031
40. Fisher E, Grallert H, Klapper M, Pfäfflin A, Schrezenmeir J, Illig T, Boeing H, Döring F. Evidence for the Thr79Met polymorphism of the ileal fatty acid binding protein (FABP6) to be associated with type 2 diabetes in obese individuals. *Mol Genet Metab*. 2009;98:400–405. doi: 10.1016/j.ymgme.2009.08.001
41. Inoue K, Wang X, Saito J, Tanino Y, Ishida T, Iwaki D, Fujita T, Kimura S, Munakata M. Plasma UGRP1 levels associate with promoter G-112A polymorphism and the severity of asthma. *Allergol Int*. 2008;57:57–64. doi: 10.2332/allergolint.O-07-493
42. Zhou Z, Zuo CL, Li XS, Ye XP, Zhang QY, Wang P, Zhang RX, Chen G, Yang JL, Chen Y, et al. Uterus globulin associated protein 1 (UGRP1) is a potential marker of progression of Graves' disease into hypothyroidism. *Mol Cell Endocrinol*. 2019;494:110492. doi: 10.1016/j.mce.2019.110492
43. Powell-Wiley TM, Baumer Y, Baah FO, Baez AS, Farmer N, Mahlobo CT, Pita MA, Potharaju KA, Tamura K, Wallen GR. Social determinants of cardiovascular disease. *Circ Res*. 2022;130:782–799. doi: 10.1161/CIRCRESAHA.121.319811

Supplemental Material

Data S1.

Supplemental Methods

Elastic Net Model after partial least squares discriminant analysis.

For proteins with VIP>1, a logistic regression model with elastic net penalty[24] on each bootstrap dataset was further considered. The elastic net penalty shrinks some protein coefficients to zero in the logistic regression model so that proteins with zero coefficients can be deemed unimportant in discriminating cases and controls, and hence may be deleted. To select the elastic net regularization parameter, cross-validation using the *cv.glmnet* function in the R package *glmnet*[44] was used, and the elastic net parameter was set to $\alpha = 0.5$. PLSDA was carried out with the R package *DiscriMiner*.

Internal Validation via Bootstrap

In the absence of external validation data, we internally validated our findings using bootstrap. We repeated each step of the model development, from candidate protein biomarker selection to protein score development. We obtained 200 bootstrap samples and we kept the proportions of cases and controls similar to the proportions in the original data. For each bootstrap dataset, a two-component PLSDA model was fit, and proteins with VIP>1 were kept. This step selected 31 proteins (compared to 29 proteins in the original [discovery] analysis). There were 27 overlapping proteins meeting this criterion in both the discovery and validation PLSDA models. As in the original analysis, logistic regression models with elastic net were fit on each bootstrap validation data for proteins with VIP>1, and proteins with frequency of nonzero coefficients in the top 25th percentile were used to develop protein scores. We identified the same 8 proteins (FAM3B, ITGA11, IL6, HGF, CCL25, GT, PLA2G7, and SCGB3A2) that were found in the discovery analysis. We obtained 200 log-odds ratios for each protein using the bootstrap validation data, and a weighted mean (over 200 log-odds ratios) was ascertained. In the discovery analysis, the protein score and models were developed using the full dataset, but in order not to use the full dataset in

the validation step, we obtained a different set of bootstrap dataset (200 bootstrap datasets), and we developed protein scores for each bootstrap dataset. Then for each bootstrap data, we fit the two models: baseline, and baseline and protein score to obtain odds ratios and AUCs and their corresponding variances. A fixed-effect meta-analysis using inverse-variance method was used to obtain weighted averages and confidence intervals. The protein score from the validation analysis was statistically significant with an odds ratio of 2.78 (CI: 2.67-2.90) compared to an odds ratio of 2.17 (CI: 1.58 - 2.99) from the discovery set. The validation AUCs for the baseline and baseline plus protein score models were respectively 0.74 (CI: 0.73-0.75) and 0.79(CI: 0.78-0.80) (Table S12), representing an improvement in AUC above the baseline model of 6.8% compared to 5.8% in the discovery analyses.

Data S2. Appendix: Study Group Authors

FIRST Study Group:

Rodger D MacArthur, Richard M Novak, Grace Peng, Li Chen, Ying Xiang, Katherine Huppler Hullsiek, Michael J Kozal, Mary van den Berg-Wolf, Christopher Henely, Barry Schmetter, Marjorie Dehlinger, for the CPCRA 058 Study Team and the Terry Beirn Community Programs for Clinical Research on AIDS (CPCRA)

Contributors: RDM, RMN, GP, LC, MBW, BS, and MD designed the study and wrote the study protocol. RDM, RMN, GP, LC, YX, MJK, MBW, CH, and BS implemented the study. GP, LC, YX, and KHH did the statistical analyses. RDM, RMN, MJK, and MBW were responsible for patient accrual. All authors assisted with interpretation of results and manuscript preparation.

Trial participants:
Southern New Jersey AIDS Clinical Trials, Camden NJ, USA (K Casey, D Condoluci, D McIntyre); Wayne State University, Detroit MI, USA (L R Crane, P Chandrasekar, M Farrough); Houston AIDS Research Team, Houston TX, USA (R C Arduino, M Rodriguez-Barradas, F Visnegarwala); Richmond AIDS Consortium, Richmond, VA, USA (E Fisher, C Clark, M Pittman); North Jersey Community Research Initiative, Newark, NJ, USA (G Perez, R Roland, N Regevik); AIDS Research Alliance, Chicago, IL, USA (J P Uy, R Luskin-Hawk, M Giles); Harlem AIDS Treatment Group, New York, NY, USA (W El-Sadr, F Siegal, R Contreras); Henry Ford Hospital, Detroit, MI, USA (N P Markowitz, I Brar, L H Makohon); Denver Community Program for Clinical Research on AIDS, Denver, CO, USA (D Cohn, F Moran, R Fernandez); Louisiana Community AIDS Research Program, New Orleans, LA, USA (D Mushatt, M O Stuart, D L Dandridge); Research and Education Group, Portland OR, USA (J Sampson, D Antoniskis, J Leggett); Wide-Reaching AIDS Program, Washington DC, USA (M Turner, C Jones, D Thomas); Community Consortium of San Francisco, San Francisco CA, USA (R C Scott, P Pell); Bronx AIDS Research Consortium, Bronx, NY, USA (E E Telzak, R B Cindrich, J Shuter); Temple University, Philadelphia, PA, USA (E M Tedaldi, K Lattanzi, S Frederick); New England Program for AIDS Clinical Trials, New Haven, CT, USA (J Jensen, L Daly, L Andrews); Partners in Research New Mexico, Albuquerque, NM, USA (S B Williams, C S Nicholson, K Hammer); AIDS Research Consortium of Atlanta, Atlanta, GA, USA (B Sweeton, L Miller, M Thompson).

ESPRIT and SILCAT Study groups:

Writing Group: Members of the writing group for the International Network for Strategic Initiatives in Global HIV Trials (INSIGHT)– Evaluation of Subcutaneous Proleukin in a Randomized International Trial (ESPRIT) Study Group and the Subcutaneous Recombinant, Human Interleukin-2 in HIV-Infected Patients with Low CD4+ Counts under Active Antiretroviral Therapy (SILCAAT) Scientific Committee (D. Abrams, M.D. [cochair], Y. Lévy, M.D. [cochair], M.H. Losso, M.D. [cochair], A. Babiker, Ph.D., G.

Collins, M.S., D.A. Cooper, M.D., J. Darbyshire, M.B., Ch.B., S. Emery, Ph.D., L. Fox, M.D., F. Gordin, M.D., H.C. Lane, M.D., J.D. Lundgren, M.D., R. Mitsuyasu, M.D., J.D. Neaton, Ph.D., A. Phillips, Ph.D., J.P. Routy, M.D., G. Tambussi, M.D., and D. Wentworth, M.P.H.

Coordinating Centers: Copenhagen: B Aagaard, E Aragon, J Arnaiz, L Borup, B Clotet, U Dragsted, A Fau, D Gey, J Grarup, U Hengge, P Herrero, P Jansson, B Jensen, K Jensen, H Juncher, P Lopez, J Lundgren, C Matthews, D Mollerup, M Pearson, A Phillips, S Reilev, K Tillmann, S Varea. London: B Angus, A Babiker, B Cordwell, J Darbyshire, W Dodds, S Fleck, J Horton, F Hudson, Y Moraes, F Pacciarini, A Palfreeman, N Paton, N Smith, F van Hooff.

Minneapolis: J Bebchuk, G Collins, E Denning, A DuChene, L Fosdick, M Harrison, K Herman-Lamin, E Krum, G Larson, J Neaton, R Nelson, K Quan, S Quan, T Schultz, G Thompson, D Wentworth, N Wyman. Sydney: C Carey, F Chan, D Cooper, B Cordwell, D Courtney-Rodgers, F Drummond, S Emery, M Harrod, S Jacoby, L Kearney, M Law, E Lin, S Pett, R Robson, N Seneviratne, M Stewart, E Watts.

Washington: E Finley, F Gordin, A Sánchez, B Standridge, M Vjecha.

Endpoint Review Committee: W Belloso, R Davey, D Duprez, J Gatell, J Hoy, A Lifson, C Pederson, G Perez, R Price, R Prineas, F Rhame, J Sampson, J Worley.

Data and Safety Monitoring Board: J Modlin, V Beral, R Chaisson, T Fleming, C Hill, K Kim, B Murray, B Pick, M Seligmann, I Weller.

National Institute of Allergy and Infectious Disease: K Cahill, L Fox, M Luzar, A Martinez, L McNay, J Pierson, J Tierney, S Vogel.

International Drug Distribution (CTS Inc., Durham, North Carolina): V Costas, J Eckstrand.

Specimen Repository (SAIC Frederick, Inc.): S Brown.

Clinical Sites for ESPRIT and/or SILCAAT: Argentina: L Abusamra, E Angel, S Aquilia, W Belloso, J Benetucci, V Bittar, E Bogdanowicz, P Cahn, A Casiro, J Contarelli, J Corral, L Daciuk, D David, W Dobrzanski, A Duran, J Ebenrstejin, I Ferrari, D Fridman, V Galache, G Guaragna, S Ivalo, A Krolewiecki, I Lanusse, H Laplume, M Lasala, R Lattes, J Lazovski, G Lopardo, M Losso, L Lourtau, S Lupo, A Maranzana, C Marson, L Massera, G Moscatello, S Olivia, I Otegui, L Palacios, A Parlante, H Salomon, M Sanchez, C Somenzini, C Suarez, M Tocci, J Toibaro, C Zala. **Australia:** S Agrawal, P Ambrose, C Anderson, J Anderson, D Baker, K Beileiter, K Blavius, M Bloch, M Boyle, D Bradford, P Britton, P Brown, T Busic, A Cain, L Carrall, S Carson, I Chenoweth, J Chuah, F Clark, J Clemons, K Clezy, D Cooper, P Cortissos, N Cunningham, M Curry, L Daly, C D'Arcy-Evans, R Del Rosario, S Dinning, P Dobson, W Donohue, N Doong, C Downs, E Edwards, S Edwards, C Egan, W Ferguson, R Finlayson, C Forsdyke, L Foy, T Franic, A Frater, M French, D Gleeson, J Gold, P Habel, K Haig, S Hardy, R Holland, J Hoy, J Hudson, R Hutchison, N Hyland, R James, C Johnston, M Kelly, M King, K Kunkel, H Lau, J Leamy, D Lester, J Leung, A Lohmeyer, K Lowe, K MacRae, C Magness, O Martinez, H Maruszak, N Medland, S Miller, J Murray, P Negus, R Newman, M Ngieng, C Nowlan, J Oddy, N

Orford, D Orth, J Patching, M Plummer, S Price, R Primrose, I Prone, H Ree, C Remington, R Richardson, S Robinson, G Rogers, J Roney, N Roth, D Russell, S Ryan, J Sarangapany, T Schmidt, K Schneider, C Shields, C Silberberg, D Shaw, J Skett, D Smith, T Meng Soo, D Sowden, A Street, B Kiem Tee, Jl Thomson, S Topaz, R Vale, C Villella, A Walker, A Watson, N Wendt, L Williams, D Youds. **Austria:** A Aichelburg, P Cichon, B Gemeinhart, A Rieger, B Schmied, V Touzeau-Romer, N Vetter. **Belgium:** R Colebunders, N Clumeck, A DeRoo, K Kabeya, E O'Doherty, S de Wit. **Brazil:** C De Salles Amorim, C Basso, S Flint, E Kallas, G Levi, D Lewi, L Pereira Jr, M da Silva, T Souza, A Toscano. **Canada (CTN):** J Angel, M Arsenault, M Bast, B Beckthold, P Bouchard, I Chabot, R Clarke, J Cohen, P Coté, M Ellis, C Gagne, J Gill, M Houde, B Johnston, N Jubinville, C Kato, N Lamoureux, G Larson, J LatendrePaquette, A Lindemulder, A McNeil, N McFarland, J Montaner, C Morrisseau, R O'Neill, G Page, A Piche, B Pongracz, H Preziosi, L Puri, A Rachlis, E Ralph, I Raymond, D Rouleau, JP Routy, R Sandre,T Seddon, S Shafran, C Sikora, F Smaill, D Stromberg, S Trottier, S Walmsley, K Weiss, K Williams, D Zarowny. **Denmark:** B Baadegaard, Å Bengaard Andersen, K Boedker, P Collins, J Gerstoft, L Jensen, H Moller, P Lehm Andersen, I Loftheim, L Mathiesen, H Nielsen, N Obel, C Pedersen, D Petersen, L Pors Jensen, F Trunk Black. **France (ANRS):** JP Aboulker, A Aouba, M Bensalem, H Berthe, C Blanc, M Bloch, D Bornarel, O Bouchaud, F Boue, E Bouvet, C Brancon, S Breaud, D Brosseau, A Brunet, C Capitant, C Ceppi, C Chakvetadze, C Cheneau, JM Chennebault, P De Truchis, AM Delavalle, JF Delfraissy, P Dellamonica, AM Delavalle, C Dumont, N Edeb, G Fabre, S Ferrando, A Foltzer, V Foubert, JA Gastaut, J Gerbe, PM Girard, C Goujard, B Hoen, P Honore, H Hue,T Hynh, C Jung, S Kahi, C Katlama, JM Lang, V Le Baut, B Lefebvre, N Leturque, Y Lévy, J Loison, G Maddi, A Maignan, C Majerholc, C de Boever, JL Meynard, C Michelet, C Michon, M Mole, E Netzer, G Pialoux, I Poizot-Martin, F Raffi, M Ratajczak, I Ravaux, J Reynes, D Salmon-Ceron, M Sebire, A Simon, L Tegna, D TisneDessus, C Tramoni, JP Viard, M Vidal, C Viet-Peaucelle, L Weiss, A Zeng, D Zucman. **Germany:** A Adam, K Arasteh, G Behrens, F Bergmann, M Bickel, D Bittner, J Bogner, N Brockmeyer, N Darrelmann, M Deja, M Doerler, S Esser, G Faetkenheuer, S Fenske, S Gajetzki, D Gey, F Goebel, D Gorriahn, E Harrer, T Harrer, H Hartl, M Hartmann, S Heesch, W Jakob, H Jäger, H Klinker, G Kremer, C Ludwig, K Mantzsch, S Mauss, A Meurer, A Niedermeier, N Pittack, A Plettenberg, A Potthoff, M Probst, M Rittweger, J Rockstroh, B Ross, J Rotty, E Rund, T Ruzicka, Rt Schmidt, G Schmutz, E Schnaitmann, D Schuster, T Sehr, B Spaeth, S Staszewski, HJ Stellbrink, C Stephan, T Stockey, A Stoehr, K Tillmann, A Trein, T Vaeth, M Vogel, J Wasmuth, C Wengenroth, R Winzer, E Wolf. **Ireland:** F Mulcahy, Dl Reidy. **Israel:** Y Cohen, G Drora, I Eliezer, O Godo, E Kedem, E Magen, M Mamorsky, S Pollack, Z Sthoeger, H Vered, I Yust. **Italy:** F Aiuti, M Bechi, A Bergamasco, D Bertelli, R Bruno, L Butini, M Cagliuso, G Carosi, S Casari, V Chrysoula, G Cologni, V Conti, A Costantini, A Corpolongo, G D'Offizi, F Gaiottino, M Di Pietro, R Esposito, G Filice, M Francesco, E Gianelli, C Graziella, L Magenta, F Martellotta, R Maserati, F Mazzotta, G Murdaca, G Nardini, S Nozza, F Puppo, M Pogliaghi, D Ripamonti, C Ronchetti, S Rusconi, V Rusconi, P Sacchi, N Silvia, F Suter, G Tambussi, A Uglietti, M Vechi, B Vergani, F Vichi, P Vitiello. **Japan:** A Iwamoto, Y Kikuchi, N Miyazaki, M Mori, T Nakamura, T Odawara, S Oka, T Shirasaka, M Tabata, M Takano, C Ueta, D Watanabe, Y Yamamoto. **Morocco:** I Erradey, H Himmich, K Marhoum El Filali. **The Netherlands:** W Blok, R van Boxtel, K Brinkman H Doevelaar, A van Eeden, M Grijsen, M Groot, J Juttmann, M Kuipers, S Ligthart, P van der Meulen, J Lange, N Langebeek, S Ligthart, P Reiss, C Richter, M Schoemaker, L Schrijnders-Gudde, E Septer-Bijleveld, H Sprenger, J Vermeulen, R ten Kate, B van de Ven. **Norway:** J Bruun, D Kvale, A Maeland. **Poland:** E Bakowska, M Beniowski, A Boron-Kaczmarska, J Gasiorowski, A Horban, M Inglot, B Knysz, E Mularska, M Parczewski, M Pynka, W

Rymer, A Szymczak. **Portugal:** M Aldir, F Antunes, C Baptista, J da Conceicao Vera, M Doroana, K Mansinho, C Raquel A dos Santos, E Valadas, I Vaz Pinto. **Singapore:** E Chia, E Foo, F Karim, PL Lim, A Panchalingam, N Paton, A Quek. **Spain:** R AlcázarCaballero, E Aragon, J Arnaiz, J Arribas, J Arrizabalaga, X de Barron, F Blanco, E Bouza, I Bravo, S Calvo, L Carbonero, I Carpena, M Castro, B Clotet, L Cortes, M del Toro, P Domingo, M Elias, J Espinosa, V Estrada, E Fernandez-Cruz, P Fernández, H Freud, M Fuster, A Garcia, G Garcia, R Garrido, J Gatell, P Gijón, J Gonzalez-García, I Gil, A González, J González-Lahoz, P López Grosso, M Gutierrez, E Guzmán, J Iribarren, M Jiménez, A Jou, J Juega, J Lopez, P Lopez, F Lozano, L Martín-Carbonero, R Mata, G Mateo, A Menasalvas, C Mirelles, J de Miguel Prieto, M Montes, A Moreno, J Moreno, V Moreno, R Muñoz, A Ocampo, E Ortega, L Ortiz, B Padilla, A Parras, A Paster, J Pedreira, J Peña, R Perea, B Portas, J Puig, F Pulido, M Rebollar, J de Rivera, V Roca, F Rodríguez- Arrondo, R Rubio, J Santos, J Sanz, G Sebastian, M Segovia, V Soriano, L Tamargo, S Varea, P Viciana, M von Wichmann. **Sweden:** G Bratt, A Hollander, P Olov Pehrson, I Petz, E Sandstrom, A Sönnerborg. **Switzerland:** E Bernasconi, V Gurtner. **Thailand:** U Ampunpong, C Auchieng, C Bowonwatanuwong, P Chanchai, P Chetchotisakd, T Chuenyan, C Duncombe, M Horsakulthai, P Kantipong, K Laohajinda, P Phanuphak, V Pongsurachet, S Pradapmook, K Ruxruntham, S Seekaew, A Sonjai, S Suwanagool, W Techasathit, S Ubolyam, J Wankoon. **United Kingdom:** I Alexander, D Dockrell, P Easterbrook, B Edwards, E Evans, M Fisher, R Fox, B Gazzard, G Gilleran, J Hand, L Heald, C Higgs, S Jebakumar, I Jendrulek, M Johnson, S Johnson, F Karim, G Kinghorn, K Kuldanek, C Leen, R Maw, S McKernan, L McLean, S Morris, M Murphy, S O'Farrell, E Ong, B Peters, C Stroud, M Wansbrough-Jones, J Weber, D White, I Williams, M Wiselka, T Yee. **United States:** S Adams, D Allegra, L Andrews, B Aneja, G Anstead, R Arduino, R Artz, J Bailowitz, S Banks, J Baxter, J Baum, D Benator, D Black, D Boh, T Bonam, M Brito, J Brockelman, S Brown, V Bruzzese, A Burnside Jr., V Cafaro, K Casey, L Cason, G Childress, Cl Clark, D Clifford, M Climo, D Cohn, P Couey, H Cuervo, R Davey Jr, S Deeks, M Dennis, M Diaz-Linares, D Dickerson, M Diez, J Di Puppo, P Dodson, D Dupre, R Elion, K Elliott, W El-Sadr, M Estes, J Fabre, M Farrough, J Flamm, S Follansbee, C Foster, C Frank, J Franz, G Frechette, G Freidland, J Frische, L Fuentes, C Funk, C Geisler, K Genther, M Giles, M Goetz, M Gonzalez, C Graeber, F Graziano, D Grice, B Hahn, C Hamilton, S Hassler, A Henson, S Hopper, M John, L Johnson, M Johnson, R Johnson, R Jones, J Kahn, M Kelly, N Klimas, M Kolber, S Koletar, A Labriola, R Larsen, F Lasseter, M Lederman, T Ling, T Lusch, R MacArthur, C Machado, L Makohon, J Mandelke, S Mannheimer, N Markowitz, M Martínez, N Martinez, M Mass, H Masur, D McGregor, D McIntyre, J McKee, D McMullen, M Mettinger, S Middleton, J Mieras, D Mildvan, P Miller, T Miller, V Mitchell, R Mitsuyasu, A Moanna, C Mogridge, F Moran, R Murphy, D Mushatt, R Nahass, D Nixon, S O'Brien, J Ojeda, P Okhuysen, M Olson, J Osterberger, W Owen, Sr. S Pablovich, S Patel, G Perez, G Pierone Jr., R Poblete, A Potter, E Preston, C Rappoport, N Regevik, M Reyelt, F. Rhame, L Riney, M Rodriguez-Barradas, M Rodriguez, Milagros Rodriguez, J Rodriguez, R Roland, C Rosmarin-DeStefano, W Rossen, J Rouff, M Saag, J Sampson, S Santiago, J Sarria, S Wirtz, U Schmidt, C Scott, A Sheridan, A Shin, S Shrader, G Simon, D Slowinski, K Smith, J Spotkov, C Sprague, D States, C Suh, J Sullivan, K Summers, B Sweeton, V Tan, T Tanner, E Tedaldi, Z Temesgen, D Thomas, M Thompson, C Tobin, N Toro, W Towner, K Upton, J Uy, S Valenti, C van der Horst, J Vita, J Voell, J Walker, T Walton, K Wason, V Watson, A Wellons, J Weise, M White, T Whitman, B Williams, N Williams, J Windham, M Witt, K Workowski, G Wortmann, T Wright, C Zelasky, B Zwickl.

SMART Study Group:

Writing Group: W.M. El-Sadr [cochair], Harlem Hospital Center and Columbia University, New York; J.D. Lundgren [cochair], Hvidovre University Hospital, Denmark; J.D. Neaton [cochair], University of Minnesota, Minneapolis; F. Gordin, Washington Veterans Affairs Medical Center, Washington, DC; D. Abrams, University of California, San Francisco; R.C. Arduino, University of Texas Medical School, Houston; A. Babiker, Medical Research Council, London; W. Burman, Denver Public Health Department; N. Clumeck, Centre Hospitalier Universitaire SaintPierre, Brussels; C.J. Cohen, Community Research Initiative of New England, Boston; D. Cohn, Denver Public Health Department; D. Cooper, National Centre in HIV Epidemiology and Clinical Research, Sydney; J. Darbyshire, Medical Research Council, London; S. Emery, National Centre in HIV Epidemiology and Clinical Research, Sydney; G. Fätkenheuer, University Hospital, Cologne; B. Gazzard, Medical Research Council, London; B. Grund, University of Minnesota, Minneapolis; J. Hoy, National Centre in HIV Epidemiology and Clinical Research, Melbourne; K. Klingman, National Institute of Allergy and Infectious Diseases, Bethesda, MD; M. Losso, Hospital General de Agudos J.M. Ramos Mejia, Buenos Aires; N. Marko witz, Henry Ford Hospital, Detroit; J. Neu haus, University of Minnesota, Minneapolis; A. Phillips, Royal Free Hospital School of Medicine, London; and C. Rappoport, University of California, San Francisco

SMART study group are as follows:
Community Programs for Clinical Research on AIDS Chair's Office and Operations Center — F. Gordin (group leader), E. Finley, D. Dietz, C. Chesson, M. Vjecha, B. Standridge, B. Schmetter, L. Grue, M. Willoughby, A. Demers;

Regional Coordinating Centers — Copenhagen — J.D. Lundgren, A. Phillips, U.B. Dragsted, K.B. Jensen, A. Fau, L. Borup, M. Pearson, P.O. Jansson, B.G. Jensen, T.L. Benfield; London — J.H. Darbyshire, A.G. Babiker, A.J. Palfreeman, S.L. Fleck, Y. Collaco-Moraes, B. Cordwell, W. Dodds, F. van Hooff, L. Wyzydrag; Sydney — D.A. Cooper, S. Emery, F.M. Drummond, S.A. Connor, C.S. Satchell, S. Gunn, S. Oka, M.A. Delfino, K. Merlin, C. McGinley;

Statistical and Data Management Center — Minneapolis — J.D. Neaton, G. Bartsch, A. DuChene, M. George, B. Grund, M. Harrison, C. Hogan (deceased), E. Krum, G. Larson, C. Miller, R. Nelson, J. Neuhaus, M.P. Roediger, T. Schultz, L. Thackeray;

Electrocardiography Reading Center — R. Prineas, C. Campbell;

End Point Review Committee — G. Perez (cochair), A. Lifson (cochair), D. Duprez, J. Hoy, C. Lahart, D. Perlman, R. Price, R. Prineas, F. Rhame, J. Sampson, J. Worley;

NIAID Data and Safety Monitoring Board — M. Rein (chair), R. Der Simonian (executive secretary), B.A. Brody, E.S. Daar, N.N. Dubler, T.R. Fleming, D.J. Freeman, J.P. Kahn, K.M. Kim, G. Medoff, J.F. Modlin, R. Moellering, Jr., B.E. Murray, B. Pick, M.L. Robb, D.O. Scharfstein, J. Sugarman, A. Tsiatis, C. Tuazon, L. Zoloth; NIAID — K. Klingman, S. Lehrman;

SMART Clinical Site Investigators (numbers of enrolled patients are in parentheses)— Argentina (147) — J. Lazovski, W.H. Belloso, M.H. Losso, J.A. Benetucci, S. Aquilia, V. Bittar, E.P. Bogdanowicz, P.E. Cahn, A.D. Casiró, I. Cassetti, J.M. Contarelli, J.A. Corral, A. Crinejo, L. Daciuk, D.O. David, G. Guaragna, M.T. Ishida, A. Krolewiecki, H.E. Laplume, M.B. Lasala, L. Lourtau, S.H. Lupo, A. Maranzana, F. Masciottra, M. Michaan, L. Ruggieri, E. Salazar, M. Sánchez, C. Somenzini; **Australia** (170) — J.F. Hoy, G.D. Rogers, A.M. Allworth, J.S.C. Anderson, J. Armishaw, K. Barnes, A. Carr, A. Chiam, J.C.P. Chuah, M.C. Curry, R.L. Dever, W.A. Donohue, N.C. Doong, D.E. Dwyer, J. Dyer, B. Eu, V.W. Ferguson, M.A.H. French, R.J. Garsia, J. Gold, J.H. Hudson, S. Jeganathan, P. Konecny, J. Leung, C.L. McCormack, M. McMurchie, N. Medland, R.J. Moore, M.B. Moussa, D. Orth, M. Piper, T. Read, J.J. Roney, N. Roth, D.R. Shaw, J. Silvers, D.J. Smith, A.C. Street, R.J. Vale, N.A. Wendt, H. Wood, D.W. Youds, J. Zillman; **Austria** (16) — A. Rieger, V. Tozeau, A. Aichelburg, N. Vetter; **Belgium** (95) — N. Clumeck, S. Dewit, A. de Roo, K. Kabeya, P. Leonard, L. Lynen, M. Moutschen, E. O'Doherty; **Brazil** (292) — L.C. Pereira, Jr., T.N.L. Souza, M. Schechter, R. Zajdenverg, M.M.T.B. Almeida, F. Araujo, F. Bahia, C. Brites, M.M. Caseiro, J. Casseb, A. Etzel, G.G. Falco, E.C.J. Filho, S.R. Flint, C.R. Gonzales, J.V.R. Madruga, L.N. Passos, T. Reuter, L.C. Sidi, A.L.C. Toscano; **Canada** (102) — D. Zarowny, E. Cherban, J. Cohen, B. Conway, C. Dufour, M. Ellis, A. Foster, D. Haase, H. Haldane, M. Houde, C. Kato, M. Klein, B. Lessard, A. Martel, C. Martel, N. McFarland, E. Paradis, A. Piche, R. Sandre, W. Schlech, S. Schmidt, F. Smaill, B. Thompson, S. Trottier, S. Vezina, S. Walmsley; **Chile** (49) — M.J. Wolff Reyes, R. Northland; Denmark (19) — L. Ostergaard, C. Pedersen, H. Nielsen, L. Hergens, I.R. Loftheim, K.B. Jensen; **Estonia** (5) — M. Raukas, K. Zilmer; **Finland** (21) — J. Justinen, M. Ristola; **France** (272) — P.M. Girard, R. Landman, S. Abel, S. Abgrall, K. Amat, L. Auperin, R. Barruet, A. Benalycherif, N. Benammar, M. Bensalem, M. Bentata, J.M. Besnier, M. Blanc, O. Bouchaud, A. Cabié, P. Chavannet, J.M. Chennebault, S. Dargere, X. de la Tribonniere, T. Debord, N. Decaux, J. Delgado, M. Dupon, J. Durant, V. Frixon-Marin, C. Genet, L. Gérard, J. Gilquin, B. Hoen, V. Jeantils, H. Kouadio, P. Leclercq, J.-D. Lelièvre, Y. Levy, C.P. Michon, P. Nau, J. Pacanowski, C. Piketty, I. Poizot-Martin, I. Raymond, D. Salmon, J.L. Schmit, M.A. Serini, A. Simon, S. Tassi, F. Touam, R. Verdon, P. Weinbreck, L. Weiss, Y. Yazdanpanah, P. Yeni; **Germany** (215) — G. Fätkenheuer, S. Staszewski, F. Bergmann, S. Bitsch, J.R. Bogner, N. Brockmeyer, S. Esser, F.D. Goebel, M. Hartmann, H. Klinker, C. Lehmann, T. Lennemann, A. Plettenberg, A. Potthof, J. Rockstroh, B. Ross, A. Stoehr, J.C. Wasmuth, K. Wiedemeyer, R. Winzer; **Greece** (95) — A. Hatzakis, G. Touloumi, A. Antoniadou, G.L. Daikos, A. Dimitrakaki, P. Gargalianos-Kakolyris, M. Giannaris, A. Karafoulidou, A. Katsambas, O. Katsarou, A.N. Kontos, T. Kordossis, M.K. Lazanas, P. Panagopoulos, G. Panos, V. Paparizos, V. Papastamopoulos, G. Petrikkos, H. Sambatakou, A. Skoutelis, N. Tsogas, G. Xylomenos; **Ireland** (2) — C.J. Bergin, B. Mooka; **Israel** (13) — S. Pollack, M.G. Mamorksy, N. Agmon-Levin, R. Karplus, E. Kedem, S. Maayan, E. Shahar, Z. Sthoeger, D. Turner, I. Yust; **Italy** (88) — G. Tambussi, V. Rusconi, C. Abeli, M. Bechi, A. Biglino, S. Bonora, L. Butini, G. Carosi, S. Casari, A. Corpolongo, M. De Gioanni, G. Di Perri, M. Di Pietro, G. D'Offizi, R. Esposito, F. Mazzotta, M. Montroni, G. Nardini, S. Nozza, T. Quirino, E. Raise; **Japan** (15) — M. Honda, M. Ishisaka; **Lithuania** (4) — S. Caplinskas, V. Uzdaviniene; **Luxembourg** (3) — J.C. Schmit, T. Staub; **Morocco** (42) — H. Himmich, K. Marhoum El Filali; **New Zealand** (7) — G.D. Mills, T. Blackmore, J.A. Masters, J. Morgan, A. Pithie; **Norway** (17) — J. Brunn, V. Ormasssen; **Peru** (57) — A. La Rosa, O. Guerra, M. Espichan, L. Gutierrez, F. Mendo, R. Salazar; **Poland** (54) — B. Knytz, A. Horban, E. Bakowska, M. Beniowski, J. Gasiorowski, J. Kwiatkowski; **Portugal** (73) — F. Antunes, R.S. Castro, M. Doroana, A. Horta, K. Mansinho, A.C. Miranda, I.V. Pinto, E. Valadas, J. Vera; **Russia** (17) — A. Rakhmanova, E. Vinogradova, A. Yakovlev, N. Zakharova; **South Africa** (26) — R.

Wood, C. Orrel; **Spain** (100) — J. Gatell, J.A. Arnaiz, R. Carrillo, B. Clotet, D. Dalmau, A. González, Q. Jordano, A. Jou, H. Knobel, M. Larrousse, R. Mata, J.S. Moreno, E. Oretaga, J.N. Pena, F. Pulido, R. Rubio, J. Sanz, P. Viciana; **Switzerland** (91) — B. Hirschel, R. Spycher, M. Battegay, E. Bernasconi, S. Bottone, M. Cavassini, A. Christen, C. Franc, H.J. Furrer, A. Gayet-Ageron, D. Genné, S. Hochstrasser, L. Magenta, C. Moens, N. Müller, R. Nüesch; **Thailand** (159) — P. Phanuphak, K. Ruxrungtham, W. Pumpradit, P. Chetchotisakd, S. Dangthongdee, S. Kiertiburanakul, V. Klinbuayaem, P. Mootsikapun, S. Nonenoy, B. Piyavong, W. Prasithsirikul, P. Raksakulkarn; **United Kingdom** (214) — B.G. Gazzard, J.G. Ainsworth, J. Anderson, B.J. Angus, T.J. Barber, M.G. Brook, C.D. Care, D.R. Chadwick, M. Chikohora, D.R. Churchill, D. Cornforth, D.H. Dockrell, P.J. Easterbrook, P.A. Fox, R. Fox, P.A. Gomez, M.M. Gompels, G.M. Harris, S. Herman, A.G.A. Jackson, S.P.R. Jebakumar, M.A. Johnson, G.R. Kinghorn, K.A. Kuldanek, N. Larbalestier, C. Leen, M. Lumsden, T. Maher, J. Mantell, R. Maw, S. McKernan, L. McLean, S. Morris, L. Muromba, C.M. Orkin, A.J. Palfreeman, B.S. Peters, T.E.A. Peto, S.D. Portsmouth, S. Rajamanoharan, A. Ronan, A. Schwenk, M.A. Slinn, C.J. Stroud, R.C. Thomas, M.H. Wansbrough-Jones, H.J. Whiles, D.J. White, E. Williams, I.G. Williams, M. Youle; **United States** (2989) — D.I. Abrams, E.A. Acosta, S. Adams, A. Adamski, L. Andrews, D. Antoniskis, D.R. Aragon, R. Arduino, R. Artz, J. Bailowitz, B.J. Barnett, C. Baroni, M. Barron, J.D. Baxter, D. Beers, M. Beilke, D. Bemenderfer, A. Bernard, C.L. Besch, M.T. Bessesen, J.T. Bethel, S. Blue, J.D. Blum, S. Boarden, R.K. Bolan, J.B. Borgman, I. Brar, B.K. Braxton, U.F. Bredeek, R. Brennan, D.E. Britt, J. Brockelman, S. Brown, V. Bruzzese, D. Bulgin-Coleman, D.E. Bullock, V. Cafaro, B. Campbell, S. Caras, J. Carroll, K.K. Casey, F. Chiang, G. Childress, R.B. Cindrich, C. Clark, M. Climo, C. Cohen, J. Coley, D.V. Condoluci, R. Contreras, J. Corser, J. Cozzolino, L.R. Crane, L. Daley, D. Dandridge, V. D'Antuono, J.G. Darcourt Rizo Patron, J.A. DeHovitz, E. DeJesus, J. DesJardin, M. Diaz-Linares, C. Dietrich, P. Dodson, E. Dolce, K. Elliott, D. Erickson, M. Estes, L.L. Faber, J. Falbo, M.J. Farrough, C.F. Farthing, P. Ferrell-Gonzalez, H. Flynn, C. Frank, M. Frank, K.F. Freeman, N. French, G. Friedland, N. Fujita, L. Gahagan, K. Genther, I. Gilson, M.B. Goetz, E. Goodwin, F. Graziano, C.K. Guity, P. Gulick, E.R. Gunderson, C.M. Hale, K. Hannah, H. Henderson, K. Hennessey, W.K. Henry, D.T. Higgins, S.L. Hodder, H.W. Horowitz, M. Howe-Pittman, J. Hubbard, R. Hudson, H. Hunter, C. Hutelmyer, M.T. Insignares, L. Jackson, L. Jenny, M. John, D.L. Johnson, G. Johnson, J. Johnson, L. Johnson, J. Kaatz, J. Kaczmarski, S. Kagan, C. Kantor, T. Kempner, K. Kieckhaus, N. Kimmel, B.M. Klaus, N. Klimas, J.R. Koeppe, J. Koirala, J. Kopka, J.R. Kostman, M.J. Kozal, A. Kumar, A. Labriola, H. Lampiris, C. Lamprecht, K.M. Lattanzi, J. Lee, J. Leggett, C. Long, A. Loquere, K. Loveless, C.J. Lucasti, R. Luskin-Hawk, M. MacVeigh, L.H. Makohon, S. Mannheimer, N.P. Markowitz, C. Marks, N. Martinez, C. Martorell, E. McFeaters, B. McGee, D.M. McIntyre, J. McKee, E. McManus, L.G. Melecio, D. Melton, S. Mercado, E. Merrifield, J.A. Mieras, M. Mogyoros, F.M. Moran, K. Murphy, D. Mushatt, S. Mutic, I. Nadeem, J.P. Nadler, R. Nahass, D. Nixon, S. O'Brien, A. Ognjan, M. O'Hearn, K. O'Keefe, P.C. Okhuysen, E. Oldfield, D. Olson, R. Orenstein, R. Ortiz, J. Osterberger, W. Owen, F. Parpart, V. Pastore-Lange, S. Paul, A. Pavlatos, D.D. Pearce, R. Pelz, G. Perez, S. Peterson, G. Pierone, Jr., D. Pitrak, S.L. Powers, H.C. Pujet, J.W. Raaum, J. Ravishankar, J. Reeder, N. Regevik, N.A. Reilly, C. Reyelt, J. Riddell IV, D. Rimland, M.L. Robinson, A.E. Rodriguez, M.C. Rodriguez-Barradas, V. Rodriguez Derouen, R. Roland, C. Rosmarin, W.L. Rossen, J.R. Rouff, J.H. Sampson, M. Sands, C. Savini, S. Schrader, M.M. Schulte, C. Scott, R. Scott, H. Seedhom, M. Sension, A. Sheble-Hall, A. Sheridan, J. Shuter, L.N. Slater, R. Slotten, D. Slowinski, M. Smith, S. Snap, D.M. States, M. Stewart, G. Stringer, J. Sullivan, K.K. Summers, K. Swanson, I.B. Sweeton, S. Szabo, E.M. Tedaldi, E.E. Telzak, Z. Temesgen, D. Thomas, M.A. Thompson, S. Thompson, C. Ting Hong Bong, C. Tobin, J. Uy, A. Vaccaro, L.M. Vasco, I. Vecino, G.K.

Verlinghieri, F. Visnegarwala, B.H. Wade, V. Watson, S.E. Weis, J.A. Weise, S. Weissman, A.M. Wilkin, L. Williams, J.H. Witter, L. Wojtusic, T.J. Wright, V. Yeh, B. Young, C. Zeana, J. Zeh; **Uruguay** (3) — E. Savio, M. Vacarezza.

START Study Group:

Writing Group: Jens D. Lundgren, M.D. Abdel G. Babiker, Ph.D. [cochair], Fred Gordin, M.D. [cochair], Sean Emery, Ph.D., Birgit Grund, Ph.D., Shweta Sharma, M.S., Anchalee Avihingsanon, M.D., David A. Cooper, M.D., Gerd Fätkenheuer, M.D., Josep M. Llibre, M.D., Jean-Michel Molina, M.D., Paula Munderi, M.D., Mauro Schechter, M.D., Robin Wood, M.D., Karin L. Klingman, M.D., Simon Collins, H. Clifford Lane, M.D., Andrew N. Phillips, Ph.D., and James D. Neaton, Ph.D. [INSIGHT PI])

In addition to writing group, the following committee members contributed to the conduct of the START trial:
Community Advisory Board: C. Rappoport (INSIGHT community liaison), P.D. Aagaard, S. Collins, G.M. Corbelli,N. Geffen, C. Kittitrakul, T. Maynard, M. Meulbroek, D. Munroe, M.S. Nsubuga, D. Peavey, S. Schwarze, M. Valdez.

Substudy Chairs: J.V. Baker, D. Duprez (arterial elasticity); A. Carr, J. Hoy (bone mineral density); M. Dolan, A.Telenti (genomics); C. Grady (informed consent); G. Matthews, J. Rockstroh (liver fibrosis progression); W.H.Belloso, J.M. Kagan (monitoring); E. Wright, B. Brew, R.W. Price, K. Robertson, L. Cysique (neurology); K.M.Kunisaki, J.E. Connett, D.E. Niewoehner (pulmonary). Endpoint Review Committee: A. Lifson (chair), W.H.Belloso, R.T. Davey Jr., D. Duprez, J.M. Gatell, J. Hoy, C. Pedersen, R.W. Price, R. Prineas, J. Worley.

Central Drug Repository and Drug Distribution: K. Brekke, S. Meger, B. Baugh, J. Eckstrand, C. Gallagher, J. Myers, J. Rooney, J. Van Wyk.

Network Laboratory Group: J. Baxter, C. Carey, A. DuChene, E.B. Finley, M. George, J. Grarup, M. Hoover, R. Pedersen, C. Russell, B. Standridge.

Specimen Repositories: E. Flowers, M. Hoover, K. Smith (Advanced BioMedical Laboratories, LLC, Cinnaminson, NJ, United States); M. McGrath, S. Silver (AIDS and Cancer Specimen Resource, University of California, San Francisco, San Francisco, CA, United States).
Wake Forest ECG Reading Center, Winston-Salem, NC, United States: E.Z. Soliman, M. Barr, C. Campbell, S. Hensley, J. Hu, L. Keasler, Y. Li, T. Taylor, Z.M. Zhang.

Division of AIDS, National Institute of Allergy and Infectious Diseases, Bethesda, MD, United States: B. Alston-Smith, E. DeCarlo, K. Klingman, M. Proschan.

Data and Safety Monitoring Board: S. Bangdiwala (chair), R. Chaisson, A.R. Fleischman, C. Hill, J. Hilton, O.H.M. Leite, V.I. Mathan, B. Pick, C. Seas, P. Suwangool, G. Thimothe, F. Venter, I. Weller, P. Yeni.

Minnesota Coordinating Center, University of Minnesota, Minneapolis, MN, United States: J.D. Neaton, K. Brekke, G. Collins, E.T. Denning, A. DuChene, N.W. Engen, M. George, B. Grund, M. Harrison, K.H. Hullsiek, L.H. Klemme, E. Krum, G. Larson, S. Meger, R. Nelson, J. Neuhaus Nordwall, K. Quan, S.F. Quan, T. Schultz, S. Sharma, G. Thompson.

International Coordinating Centers: Copenhagen HIV Programme, Rigshospitalet, University of Copenhagen, Denmark: J.D. Lundgren, B. Aagaard, A.H.D. Borges, M. Eid, J. Grarup, P. Jansson, Z. Joensen, B. Nielsen, M. Pearson, R. Pedersen, A.N. Phillips; The Kirby Institute, University of New South Wales, Sydney, Australia: S. Emery, N. Berthon-Jones, C. Carey, L. Cassar, M. Clewett, D. Courtney-Rodgers, P. Findlay, S. Hough, S. Jacoby, J. Levitt, S.L. Pett, R. Robson, V. Shahamat, A. Shambrook; Medical Research Council Clinical Trials Unit at UCL, London, United Kingdom: A.G. Babiker, B. Angus, A. Arenas-Pinto, R. Bennett, N. Braimah, E. Dennis, N. Doyle, M. Gabriel, F. Hudson, B. Jackson, A. Palfreeman, N. Paton, C. Purvis, C. Russell; Veterans Affairs Medical Center, Washington, DC, United States: F. Gordin, D. Conwell, H. Elvis, E.B. Finley, V. Kan, L. Lynch, J. Royal, A. Sánchez, B. Standridge, D. Thomas, M. Turner, M.J. Vjecha.

The following investigators participated in the START study, listed by country (country lead, numbers of participants enrolled) and clinical site:
Argentina (M.H. Losso, n=216): CAICI (Instituto Centralizado de Assistencia e Investigación Clínica Integral), Rosario Santa Fe: S. Lupo, L. Marconi, D. Aguila; FUNCEI, Buenos Aires: G. Lopardo, E. Bissio, D. Fridman; Fundación IDEAA, Buenos Aires: H. Mingrone, E. Loiza, V. Mingrone; Hospital General de Agudos JM Ramos Mejia, Buenos Aires: M. Losso, J.M. Bruguera, P. Burgoa; Hospital Interzonal General de Agudos Dr. Diego Paroissien, Buenos Aires: E. Warley, S. Tavella; Hospital Italiano de Buenos Aires, Buenos Aires: W. Belloso, M. Sanchez; Hospital Nacional Profesor Alejandro Posadas, Buenos Aires: H. Laplumé, L. Daciuk; Hospital Rawson, Cordoba: D. David, A. Crinejo; Argentinean SCC, Fundación IBIS, Buenos Aires: G. Rodriguez-Loria, L. Doldan, A. Moricz, I. Otegui, I. Lanusse. **Australia** (J. Hoy, n=109): Burwood Road General Practice, Burwood, VIC: N. Doong, S. Hewitt; Centre Clinic, St Kilda, VIC: B.K. Tee; East Sydney Doctors, Darlinghurst, NSW: D. Baker, E. Odgers; Holdsworth House Medical Practice, Darlinghurst, NSW: S. Agrawal, M. Bloch; Melbourne Sexual Health Centre, Carlton, VIC: T.R.H. Read, S.J. Kent; Prahran Market Clinic, Prahran, VIC: H. Lau, N. Roth; Royal Adelaide Hospital, Adelaide, SA: L. Daly, D. Shaw; Royal Perth Hospital, Perth, WA: M. French, J. Robinson; Sexual Health & HIV Service - Clinic 2, Brisbane, QLD: M. Kelly, D. Rowling; St Vincent's Hospital, Fitzroy, VIC: D.A. Cooper, A. J. Kelleher; Taylor Square Private Clinic, Surry Hills, NSW: C. Pell, S. Dinning; The Alfred Hospital, Melbourne, VIC: J. Hoy, J. Costa; Westmead Hospital, Westmead, NSW: D.E. Dwyer, P. King. **Austria** (A. Rieger, n=7): Otto-Wagner-Spital SMZ /Baumgartner Hoehe, Vienna: N. Vetter; B. Schmied; University Vienna General Hospital, Vienna: A. Rieger, V.R. Touzeau.**Belgium** (S. de Witt, n=102): Centre Hospitalier Universitaire St. Pierre (C.H.U. St. Pierre), Brussels: S. de Witt, N. Clumeck, K. Kabeya; Institute of Tropical Medicine, Antwerp: E. Cleve, E. Florence, L. van Petersen; Universitair Ziekenhuis Gasthuisberg, Leuven: H. Ceunen, E.H. van Wijngaerden; Universitaire Ziekenhuizen Gent, Gent: T. James, L. Vandekerckhove. **Brazil** (L.C. Pereira Jr., M. Schechter, n=619): Ambulatório de Imunodeficiências (LIM-56), Sao Paulo, SP: J. Casseb, E. Constantinov, M.A. Monteiro; Center for ID at UFES, Vitoria, ES: L.N. Passos, T. Reuter; Centro de Referência e Treinamento DST/AIDS, Sao Paulo, SP: S.T. Leme, J.V.R. Madruga, R.S. Nogueira;

Hospital Escola Sao Francisco de Assis, Rio de Janeiro, RJ: M. Barbosa Souza, C. Beppu Yoshida, M. Dias Costa; Instituto de Infectologia Emilio Ribas, Sao Paulo, SP: R. Castro, R.Cruz, S. Ito, T.N. Lobato Souza; Instituto FIOCRUZ, Rio de Janeiro, RJ: B. Grinsztejn, V.G. Veloso, S. Wagner Cardoso; SEI Serviços Especializados em Infectología LTDA, Salvador, Bahia: F. Bahia, C. Brites, J. Correia. **Chile** (M.J. Wolff, n=76): Fundación Arriarán, Santiago: M. Wolff, R. Northland, C. Cortés. **Czech Republic** (D. Sedlacek, n=13): Faculty Hospital Na Bulovce, Prague: D. Jilich; University Hospital Plzen, Plzen: D. Sedlacek. **Denmark** (J. Gerstoft, n=33): Hvidovre University Hospital, Hvidovre: P. Collins, L. Mathiesen; Odense University Hospital, Odense: L. Hergens, C. Pedersen; Rigshospitalet, Copenhagen: J. Gerstoft, L.P. Jensen; Århus Universitetshospital, Skejby, Århus: I.R. Lofthiem, L. Østergaard. **Estonia** (K. Zilmer, n=8): West Tallinn Central Hospital Infectious Diseases, Tallinn: K. Zilmer. **Finland** (M. Ristola, n=23): Helsinki University Central Hospital, Helsinki: M. Ristola, O. Debnam. **France** (B. Hoen, n=111): CHU Côte de Nacre – Caen, Caen: R. Verdon, S. Dargere; CHU de Besançon -Hôpital Jean-Minjoz, Besancon: B. Hoen, C. Chirouze; Groupe Hospitalier Pitié-Salpêtrière, Paris: C. Katlama, M-A. Valantin; Hôpital Antoine Béclère, Clamart: F. Boue, I. Kansau; Hôpital de Bicêtre, Le Kremlin-Bicetre: C. Goujard, C. Chakvetadze; Hôpital Européen Georges Pompidou, Paris: L. Weiss, M Karmochkine; Hôpital Foch, Suresnes: D. Zucman, C. Majerholc; Hôpital Gustava Dron, Tourcoing: O.Robineau, R. Biekre; Hôpital Henri Mondor, Creteil: Y. Levy, J.D. Lelievre; Hôpital Hôtel Dieu, Paris: J.P. Viard, J Ghosn; Hôpital Saint-Antoine, Paris: J. Pacanowski, B. Lefebvre; Hôpital Saint-Louis, Paris: J.-M. Molina, L. Niedbalski, M. Previlon; French SCC, ANRS-Inserm SC10, Paris: J.P. Aboulker, C. Capitant, B. Lebas, N. Leturque, L. Meyer, E. Netzer. **Germany** (G. Fätkenheuer, n=312): EPIMED, Berlin: K. Arastéh, T. Meier; Gemeinschaftspraxis Jessen-Jessen-Stein, Berlin: C. Zedlack, H. Jessen; ICH Study Center, Hamburg: S. Heesch, C. Hoffmann; Ifi - Studien und Projekte GmbH, Hamburg: A. Plettenberg, A. Stoehr; Johann Wolfgang Goethe - University Hospital, Frankfurt: G. Sarrach, C. Stephan; Klinik I für Innere Medizin der Universität zu Köln, Cologne: G. Fätkenheuer, E. Thomas; Klinikum der Universität München, Munich: J.R. Bogner, I. Ott; Klinikum Dortmund GmbH, Dortmund: M. Hower, C. Bachmann; Medizinische Hochschule Hannover, Hannover: M. Stoll, R. Bieder; Medizinische Universitätsklinik - Bonn, Bonn: J. Rockstroh, B. Becker; Universitätsklinikum Düsseldorf, Düsseldorf: B. Jensen, C. Feind; Universitätsklinikum Erlangen, Erlangen: E. Harrer, T. Harrer; Universitätsklinikum Essen, Essen: S. Esser, H. Wiehler; Universitätsklinikum Heidelberg, Heidelberg: M. Hartmann, R. Röger; Universitätsklinikum Regensburg, Regensburg: B. Salzberger, E. Jäger; Universitätsklinikum Würzburg, Würzburg: H. Klinker, G. Mark; Universitätsklinikum, Hamburg-Eppendorf: J. van Lunzen, N. Zerche; German SSC, Johann Wolfgang Goethe - University Hospital, Frankfurt: V. Müller, K. Tillman. **Greece** (G. Touloumi, n=101): AHEPA University Hospital, Thessaloniki Central Macedonia: S. Metallidis, O. Tsachouridou; Attikon University General Hospital, Athens: A. Papadopoulos, K. Protopapas; Evangelismos General Hospital, Athens: A. Skoutelis, V. Papastamopoulos; Hippokration University General Hospital of Athens, Athens: H. Sambatakou, I. Mariolis; Korgialenio-Benakio Hellenic Red Cross, Athens: M. K. Lazanas, M. Chini; Syngros Hospital, Athens: S. Kourkounti, V. Paparizos; Greek SCC, National Kapodistrian University of Athens, Athens: G. Touloumi, V. Gioukari, O. Anagnostou. **India** (n=91): Institute of Infectious Diseases, Pune Maharashtra: A. Chitalikar, S. Pujari; YRGCARE Medical Centre VHS, Chennai CRS: F. Beulah, N. Kumarasamy, S. Poongulali. **Ireland** (P. Mallon, n=7): Mater Misericordiae University Hospital, Dublin: P. Mallon, P. McGettrick. **Israel** (E. Kedem, n=28): Rambam Medical Center, Haifa: E. Kedem, S. Pollack; Tel Aviv Sourasky Medical Center, Tel Aviv: D. Turner. **Italy** (G. Tambussi, n=33): Lazzaro Spallanzani IRCSS, Rome: A. Antinori, R. Libertone; Ospedale San Raffaele S.r.l., Milan: G. Tambussi, S. Nozza, M.R. Parisi. **Luxembourg** (T.

Staub, n=5): Centre Hospitalier de Luxembourg, Luxembourg: T. Staub, C. Lieunard. **Malaysia** (n=18): University Malaya Medical Centre, Kuala Lumpur: R.I.S.R. Azwa. **Mali** (S. Dao, n=41): SEREFO/ CESAC Mali, Bamako, Bamako: B. Baya, M. Cissé, D. Goita. **Mexico** (n=48): INCMNSZ (Instituto Nacional de Ciencias Médicas y Nutrición), Tlalpan D.F.: J. Sierra-Madero, M.E. Zghaib. **Morocco** (K.M. El Filali, n=44): University Hospital Centre Ibn Rochd, Casablanca: K.M. El Filali, I. Erradey, H.Himmich. **Nigeria** (n=50): Institute of Human Virology-Nigeria (IHVN), Garki, Abuja FCT: E. Ekong, N. Eriobu. **Norway** (V. Ormaasen, n=15): Oslo University Hospital, Ulleval, Oslo: V. Ormaasen, L. Skeie. **Peru** (A. La Rosa, n=215): Hospital Nacional Edgardo Rebagliati Martins, Lima, Lima: M. Espichan Gambirazzio, F. Mendo Urbina; Hospital Nacional Guillermo Almenara Irigoyen, Lima, Lima: R. Salazar Castro, J. Vega Bazalar; IMPACTA Salud y Educación, Lima, Lima: M.E. Guevara, R. Infante, J. Sanchez, M. Sanchez; IMPACTA San Miguel, Lima, Lima: R. Chinchay, J.R. Lama, M. Sanchez; Via Libre, Lima, Lima: E.C. Agurto, R. Ayarza, J.A. Hidalgo. **Poland** (A.J. Horban, n=68): EMC Instytut Medyczny SA, Wroclaw: B. Knysz, A. Szymczak; Uniwersytecki Szpital Kliniczny, Bialystok: R. Flisiak, A. Grzeszczuk; Wojewodzki Szpital Zakazny, Warsaw: A.J. Horban, E. Bakowska, A. Ignatowska. **Portugal** (L. Caldeira, n=67): Hospital Curry Cabral, Lisbon: F. Maltez, S. Lino; Hospital de Egas Moniz, Lisbon: K. Mansinho, T. Bapista; Hospital de Santa Maria, Lisbon: M. Doroana, A. Sequeira, L. Caldeira; Hospital Joaquim Urbano, Oporto: J. Mendez, R.S.E. Castro. **South Africa** (R. Wood, n=518): 1 Military Hospital, Pretoria Gauteng: S.A. Pitsi; Desmond Tutu HIV Centre - Cape Town, Cape Town, Western Province: R. Kaplan, N. Killa, C. Orrell, M. Rattley; Durban International Clinical Research Site, Durban, KwaZulu Natal: U.G. Lalloo, R. Mngqibisa, S. Pillay; Durban International Clinical Research Site WWH, Durban, KwaZulu Natal: J. Govender, M. John; University of Witwatersrand, Johannesburg, Gauteng: S. Badal-Faesen, N. Mwelase, M. Rassool. **Spain** (J.R. Arribas, n=234): Complejo Hospitalario Xeral Cies, Vigo Pontevedra: A.O. Hermida, F. Warncke; Hospital Clínic de Barcelona, Barcelona: J.M. Gatell, A. Gonzalez; Hospital Clínico San Carlos, Madrid: V. Estrada, M. Rodrigo; Hospital de la Santa Creu i Sant Pau, Barcelona: P. Domingo, M. Gutierrez; Hospital del Mar, Barcelona: H.J. Knobel, A. Gonzalez; Hospital La Paz, Madrid: J.R. Arribas, M. Montes Ramirez; Hospital La Princesa, Internal Medicine and Infectious Disease Service CRS, Madrid: I. de los Santo Gil, J. Sanz Sanz; Hospital Universitari Germans Trias i Pujol, Badalona: B. Clotet, J.M. Llibre, P. Cobarsi; Hospital Universitari Mutua Terrassa, Terrassa Barcelona: D. Dalmau, C. Badia; Hospital Universitario Doce de Octubre, Madrid: R. Rubio, M.M. del Amo; Hospital Universitario Príncipe de Asturias, Alcala de Henares Madrid: J. Sanz Moreno; Hospital Universitario y Politécnico La Fe, Valencia: J. López Aldeguer, S. Cuellar; Spanish SSC, Acoiba, Madrid: P. López, B. Portas, P. Herrero. **Sweden** (M. Gisslén, n=2): Sahlgrenska University Hospital, Sweden: M. Gisslén, L. Johansson; Skåne University Hospital, Malmö: C. Håkangård, K. Törqvist. **Switzerland** (H. Furrer, n=31): Bern University Hospital, Bern: H. Furrer, A. Rauch; Unite VIH/SIDA Genèva, Genèva: A.L. Calmy, B. Hirschel (retd), T Lecompte; University Hospital Basel, Basel: M. Stoeckle; University Hospital Zurich, Zürich: N. Muller, M. Rizo-Oberholzer; Swiss SCC, Bern University Hospital, Bern: H. Furrer, C. Bruelisauer, A. Christen, M. Lacalamita. **Thailand** (K. Ruxrungtham, n=248): Bamrasnaradura Infections Diseases Institute, Nonthaburi: W. Prasithsirikul, S. Thongyen; Chiangrai Prachanukroh Hospital, Chiang Rai: P. Kantipong, S. Khusuwan; Chonburi Regional Hospital, Chonburi: C. Bowonwatanuwong, U. Ampunpong; Chulalongkorn University Hospital, Bangkok: K. Ruxrungtham, A. Avihingsanon, W. Thiansanguankul; Khon Kaen University, Srinagarind Hospital, Khon Kaen: P. Chetchotisakd, P. Motsikapun, S. Anunnatsari; Ramathibodi Hospital, Bangkok: S. Kiertiburanakul, N. Sanmeema; Research Institute for Health Sciences (RIHES), Chiang Mai: K. Supparatpinyo, P. Sugandhavesa; Sanpatong Hospital, Chiang

Mai: V. Klinbuayaem, Y. Siriwarothai; Siriraj Hospital, Bangkok Noi: W. Ratanasuwan, T Anekthananon; Thai SCC, The HIV Netherlands Australia Thailand Research Collaboration (HIV-NAT), Bangkok: W. Harnnapachewin, T. Jupimai, P. Rerksirikul. **Uganda** (P. Mugyenyi, n=349): Joint Clinical Research Center (JCRC), Kampala: P. Mugyenyi, C. Kityo, H. Mugerwa; MRC/UVRI Research Unit on AIDS, Entebbe: P. Munderi, B. Kikaire, J. Lutaakome; MRC/UVRI Research Unit on AIDS, Masaka – satellite site: Z. Anywaine. **United Kingdom** (M.A. Johnson, n=339): Barts Health NHS Trust, London: C. Orkin, J. Hand; Belfast Health and Social Care Trust (RVH), Belfast Northern Ireland: C. Emerson, S. McKernan; Birmingham Heartlands Hospital, Birmingham West Midlands: D. White, C. Stretton; Brighton and Sussex University Hospitals NHS Trust, Brighton East Sussex: M. Fisher, A. Clarke, A. Bexley; Chelsea and Westminster Hospital, London: B. Gazzard, C. Higgs, A. Jackson; Coventry and Warwickshire NHS partnership Trust, Coventry West Midlands: S. Das, A. Sahota; Gloucestershire Royal Hospital, Gloucester: A. de Burgh-Thomas, I. Karunaratne; Guy's and St.Thomas' NHS Foundation Trust, London: J. Fox, J.M. Tiraboschi; Imperial College Healthcare NHS Trust, London: A. Winston, B. Mora-Peris; Leicester Royal Infirmary, Leicester Leicestershire: M.J. Wiselka, L. Mashonganyika; Lewisham and Greenwich NHS Trust, London: S. Kegg, T. Moussaoui; North Manchester General Hospital, Manchester: E. Wilkins, Y. Clowes; Queen Elizabeth Hospital Birmingham, Birmingham West Midlands: J. Ross, J. Harding; Royal Berkshire Hospital, Reading Berkshire: F. Chen, S. Lynch; Royal Bournemouth Hospital, Bournemouth Dorset: E. Herieka, J. Ablorde; Royal Free London NHS Foundation Trust, London: M.A. Johnson, M. Tyrer, M. Youle; Sheffield Teaching Hospital NHS Foundation Trust, Sheffield South Yorkshire: D. Dockrell, C. Bowman; Southmead Hospital, Bristol: M. Gompels, L. Jennings; St. George's Healthcare NHS Trust, London: P. Hay, O. Okolo; The James Cook University Hospital, Middlesbrough Cleveland: D.R. Chadwick, P. Lambert; University College London Medical School, London: I. Williams, A. Ashraf. **United States** (K. Henry, n=507): Adult Clinical Research Center, Newark, NJ: M. Paez-Quinde, S. Swaminathan; Boston University Medical Center, Boston, MA: I. Bica, M. Sullivan; Bronx-Lebanon Hospital Center, Bronx, NY: R.B. Cindrich, L.M. Vasco; Community Research Initiative of New England, Boston, MA: J. Green, H.B. Olivet; Cooper University Hospital, Camden, NJ: J. Baxter, Y. Smith; Cornell CRS, New York, NY: V. Hughes, T. Wilkin; Denver Public Health, Denver, CO: E.M. Gardner, J. Scott; Duke University, Durham, NC: J. Granholm, N. Thielman; Florida Department of Health in Orange/Sunshine Care Center, Orlando, FL: W.M. Carter, N.D. Desai; George Washington University Medical Center, Washington, DC: D.M. Parenti, G.L. Simon; Georgetown University Medical Center, Washington, DC: P. Kumar, M. Menna; Hennepin County Medical Center, Minneapolis, MN: J. Baker, R. Givot; Henry Ford Hospital, Detroit, MI: L.H. Makohon, N.P. Markowitz; Hillsborough County Health Department, Tampa, FL: M. Chow, C. Somboonwit; Infectious Disease Associates of Northwest Florida, Pensacola, FL: A.B. Brown, B.H. Wade; Lurie Children's Hospital, Chicago, IL: J. Jensen, A. Talsky; Maternal, Child and Adolescent Center for ID/Virology USC, Alhambra, CA: A. Kovacs, L. Spencer; Mayo Clinic, Rochester, MN: S. Rizza, Z. Temesgen; Medical College of Wisconsin, Milwaukee, WI: M. Frank, S. Parker; Montefiore Medical Center, Bronx, NY: C. Rosario, J. Shuter; Mt Sinai Hospital, Chicago, IL: K. Rohit, R. Yogev; National Military Medical Center, Bethesda, MD: I. Barahona, A. Ganesan; Naval Medical Center Portsmouth NMCP, Portsmouth, VA: S. Banks, T. Lalani; Naval Medical Center San Diego NMCSD, San Diego,CA: M.F. Bavaro, S. Echols; NICE, Southfield, MI: M. Farrough, R.D. MacArthur; NIH, Bethesda, MD: R.T.Davey Jr., R. McConnell; Ohio State University, Columbus, OH: H. Harber, S.L. Koletar; Orlando Immunology Center, Orlando, FL: E. DeJesus, A.F. Garcia; Regional Center for Infectious Disease, Greensboro, NC: K. Epperson, C.N. Van Dam; San Antonio Military Health System,

JBSA Fort Sam Houston, TX: J.F. Okulicz, T.J. Sjoberg; San Juan Hospital, San Juan, PR: M. Acevedo, L. Angeli; St. Jude Children's Research Hospital, Memphis, TN: P.M. Flynn, N. Patel; Temple University, Philadelphia, PA: C. Geisler, E. Tedaldi; Texas Children's Hospital- Baylor College of Medicine, Houston, TX: C. McMullen-Jackson, W.T. Shearer; The Research & Education Group, Portland, OR: M.D. Murphy, S.M. Sweek; Tulane University Health Sciences Center, New Orleans, LA: D. Mushatt, C. Scott; UCLA CARE 4 Families, Los Angeles, CA: M. Carter, J. Deville; UCSD Mother-Child-Adolescent HIV Program, San Diego, CA: S.A. Spector, L. Stangl; University of Florida, Department of Pediatrics, Jacksonville, FL: M.H. Rathore, K. Thoma; University of Florida, Jacksonville, FL: M. Sands, N. Wilson; University of Illinois at Chicago, Chicago, IL: R.M. Novak, T. Pearson; University of Miami, Miami, FL: M.A. Kolber, T. Tanner; University of North Carolina, Chapel Hill, NC: M. Chicurel-Bayard, E. Hoffman; University of North Texas Health Science Center, Fort Worth, TX: I. Vecino, S.E. Weis; University of Puerto Rico, San Juan, PR: I. Boneta, J. Santana; University of Texas Southwestern Medical Center, Dallas, TX: M.K. Jain, M. Santos; Veterans Affairs Greater LA Healthcare System, Los Angeles, CA: M.B. Goetz, W.L. Rossen; Virginia Commonwealth University, Richmond, VA: D. Nixon, V. Watson; Wake County Human Services, Raleigh, NC: D. Currin, C. Kronk; Wake Forest University Health Sciences, Winston-Salem, NC: L. Mosley, A. Wilkin; Washington DC Veterans Administration, Washington, DC: A.M. Labriola, D.W. Thomas; Yale University School of Medicine, New Haven, CT: D. Chodkowski, G. Friedland.

Table S1: Cause of CVD case status:

Cause of CVD	N
MI	26
Coronary Revascularization	64
Stroke	24
CVD Death or Unwitnessed Death and cause unknown	17
CVD Total	131

Table S2: Average variable importance in projection (VIP) for all 107 proteins. Some proteins were represented in more than one panel.

Protein	Average VIP	Protein	Average VIP	Protein	Average VIP	Protein	Average VIP	Protein	Average VIP
GT	2.24	PLA2G7	1.12	CCL19	0.94	CST5	0.84	PCSK9	0.79
ITGA11	1.43	EPHB4	1.11	KIM1	0.93	CCL3	0.83	COL18A1	0.79
ICAM1	1.39	CDCP1	1.11	IL10RB	0.93	LAMP3	0.83	REG3A	0.79
BTN3A2	1.30	TNFR1	1.06	TRAILR2	0.92	PGF	0.82	CPA1	0.79
SCGB3A2	1.28	IL6	1.06	MCP1	0.92	HNMT	0.82	PRSS2	0.78
TNFR2	1.28	UPAR	1.05	PRSS8	0.91	GAL4	0.82	KRT19	0.77
TNFRSF10A	1.28	CST3	1.04	CKAP4	0.90	IL8	0.82	ADM	0.77
CD83	1.28	CCL25	1.02	CDH1	0.89	TFF3	0.82	CCL11	0.77
GAS6	1.23	TCN2	1.01	CCL18	0.89	IL6	0.82	CCL20	0.76
EFEMP1	1.22	CHI3L1	1.00	SPON2	0.89	CCL14	0.810	MCP3	0.76
HOSCAR	1.20	UPA	1.00	MCP1	0.88	MMP10	0.81	CCL11	0.76
LILRB4	1.18	IL18R1	0.99	NTPROBNP	0.88	CXCL16	0.81	ST2	0.76
FAM3B	1.17	CX3CL1	0.97	LIFR	0.87	TNFSF13B	0.81	CPB1	0.75
GDF15	1.17	CTSZ	0.97	IL2RA	0.86	CTSL1	0.81	AREG	0.75

HGF	1.16	IL6	0.97	GAL9	0.85	PON3	0.81	SPON1	0.74
TNFRSF9	1.16	NCR1	0.96	IL17C	0.85	MMP7	0.80	AMBP	0.74
GRN	1.16	VEGFA	0.95	IGLC2	0.84	CLEC6A	0.80	TIMD4	0.74
TIMP1	1.15	IGFBP7	0.95	FLT3L	0.84	FGF21	0.80	BNP	0.73
IL18BP	1.13	LTBR	0.94	FGF23	0.84	ACE2	0.80	SIT1	0.72
CSF1	1.12	OPG	0.94	OPG	0.84	CLEC7A	0.79	RARRES2	0.72
								QPCT	0.71
								FGF21	0.71
								CLEC4D	0.68
								CHIT1	0.68
								CLEC4C	0.66
								ITM2A	0.66
								DPP10	0.63

Table S3: Summary statistics for the number of times elastic net model had non-zero coefficients across the 200 bootstrap sets.

Protein	Counts	Proportions
HGF	199	0.995
IL6	193	0.965
FAM3B	187	0.935
SCGB3A2	187	0.935
GT	186	0.930
CCL25	180	0.900
ITGA11	179	0.895
PLA2G7	167	0.835
GDF15	159	0.795
ICAM1	146	0.730
TNFR1	140	0.700
GRN	138	0.690
EFEMP1	128	0.640
UPAR	127	0.635
CDCP1	123	0.615
CST3	123	0.615
LILRB4	119	0.595
CSF1	119	0.595
BTN3A2	116	0.580
TCN2	116	0.580
TIMP1	115	0.575
CD83	114	0.570
GAS6	114	0.570
IL18BP	113	0.565
TNFRSF9	112	0.560
HOSCAR	112	0.560
TNFRSF10A	111	0.555
TNFR2	109	0.545
EPHB4	109	0.545

Table S4: Conditional logistic regression model of CVD on standardized protein score. (This was used to generate forest plot in Figure 1. Here, n=375 as there were 15 samples with missing values.) OR for similar model with viral load categorized as: viral load (>=500 copies/ML) vs viral load (< 500 copies/ML) was 1.31 (CI: 0.60-2.84).

Variable	Odds Ratio	Std Error	Pr(>|z|)	LCI	UCI
CD4	1.00	0.00	0.95	1.00	1.00
RNA	1.00	0.00	0.85	1.00	1.00
Females vs Males	0.53	0.47	0.18	0.21	1.33
Age	1.20	0.05	<0.01	1.08	1.33
BMI	1.01	0.03	0.63	0.96	1.07
Diabetes	0.72	0.56	0.56	0.24	2.16
Black Race	1.19	0.40	0.67	0.54	2.59
Lipid-lowering medication	0.96	0.38	0.92	0.45	2.04
BP-lowering medication	1.46	0.39	0.33	0.68	3.15
CVD	3.22	0.71	0.10	0.81	12.90
Protein Score	2.17	0.16	<0.01	1.58	2.99

Table S5: Odds ratios and AUC from categorizing protein score

Protein Score	**Case individuals (n)**	**Control individuals (n)**	**Odds ratio (95% CI)**	**AUC**
Median				0.74
Below	38	157	1.00	
Above	93	102	3.15 (1.83- 5.41)	
Top 25% vs bottom 75%				0.69
1 (Top)	55	43	2.91 (1.65,5.12)	
2	76	216	1.00	
Above 0 vs below 0 protein score				0.75
Below	39	159	1.00	
Above	92	100	3.20 (1.86 to 5.49)	

Table S6: Incremental contribution of proteins and protein score to CVD risk when added to the baseline model. The baseline model includes the coagulation biomarker D-dimer. Here, n=364.

Model	AUC	LCI	UCI	Change in AUC from Baseline	NRI among cases	NRI among controls	Overall NRI
Baseline*	0.69	0.60	0.78	-	-	-	-
+ FAM3B	0.73	0.65	0.81	0.04	0.15(-0.03,0.32)	0.02(-0.10,0.15)	0.17(-0.04,0.39)
+ ITGA11	0.70	0.61	0.79	0.01	0.08(-0.09,0.26)	-0.01(-0.13,0.12)	0.07(-0.14,0.29)
+ IL6	0.70	0.61	0.78	0.01	0.15(-0.03,0.32)	0.12(-0.01,0.24)	0.26(0.05,0.48)
+ HGF	0.73	0.64	0.81	0.04	0.23(0.06,0.40)	0.11(-0.02,0.23)	0.34(0.12,0.55)
+ CCL25	0.69	0.60	0.78	0.00	0.02(-0.16,0.19)	0.15(0.02,0.27)	0.17(-0.05,0.38)
+ GT	0.73	0.65	0.82	0.05	0.28(0.11,0.45)	0.15(0.02,0.27)	0.43(0.22,0.64)
+ PLA2G7	0.72	0.64	0.80	0.03	0.16(-0.01,0.34)	0.12(0.00,0.25)	0.29(0.07,0.50)
+SCGB3A2	0.71	0.63	0.80	0.03	0.10(-0.08,0.27)	0.12(-0.01,0.24)	0.21(0.00,0.43)
+ Protein score from all 8 proteins	0.73	0.65	0.82	0.05	0.36 (0.20, 0.53)	0.24 (0.12,0.36)	0.60 (0.40,0.81)

Baseline* is a model with the following variables: CD4, RNA, sex, age, BMI, diabetes status at baseline, prior history of CVD, lipid lowering medication, blood pressure lowering medication and D-dimer (log2 transformed).

The proteins FAM3B, IL6, HGF, GT, and PLA2G7 were each statistically significant (p-value<0.05) when added to the baseline model.

Protein score was statistically significant ($p < 0.00001$) when added to the baseline model. Odds ratio for protein score: 2.09 (CI: 1.52, 2.88)

Of note, D-dimer was statistically significant (p-value =0.03) in the baseline model but not in the baseline + protein score model (p-value>0.05).

Integrated Discriminant Index when protein score was added to the baseline* model was 0.08 (0.05, 0.10).

Table S7: Incremental contribution of proteins and protein score to CVD risk when added to the baseline model. The baseline model includes the inflammatory biomarker hsCRP. Here, n=364.

Model	AUC	LCI	UCI	Change in AUC from Baseline	NRI among cases	NRI among controls	Overall NRI
Baseline*	0.73	0.65	0.81	-	-	-	-
+ FAM3B	0.72	0.63	0.80	-0.01	0.26(0.09,0.43)	0.02(-0.10,0.15)	0.29(0.07,0.50)
+ ITGA11	0.73	0.64	0.81	0.00	0.08(-0.09,0.26)	0.01(-0.12,0.13)	0.09(-0.13,0.31)
+ IL6	0.70	0.62	0.79	-0.03	0.03(-0.14,0.21)	0.13(0.01,0.26)	0.17(-0.05,0.38)
+ HGF	0.72	0.64	0.80	-0.01	0.21(0.04,0.39)	0.12(-0.01,0.24)	0.33(0.11,0.54)
+ CCL25	0.73	0.64	0.81	-0.01	0.05(-0.13,0.23)	0.12(0.00,0.25)	0.17(-0.04,0.39)
+ GT	0.73	0.65	0.81	0.00	0.30(0.13,0.46)	0.16(0.03,0.28)	0.45(0.24,0.66)
+ PLA2G7	0.72	0.63	0.80	-0.01	0.16(-0.01,0.34)	0.14(0.02,0.27)	0.30(0.09,0.52)
+SCGB3A2	0.72	0.62	0.79	-0.02	0.13(-0.04,0.31)	0.12(0.00,0.25)	0.26(0.04,0.47)
+ Protein score from all 8 proteins	0.75	0.66	0.83	0.02	0.37 (0.20, 0.53)	0.19 (0.07,0.31)	0.56 (0.35,0.76)

Baseline* is a model with the following variables: CD4, RNA, sex, age, BMI, diabetes status at baseline, prior history of CVD, lipid lowering medication, blood pressure lowering medication and hsCRP (log2 transformed).

The proteins FAM3B, IL6, HGF, GT, and PLA2G7 were each statistically significant (p-value<0.05) when added to the baseline model. SCGB3A2 was marginally significant (p-value=0.051).

Protein score was statistically significant ($p < 0.00001$) when added to the baseline model. Odds ratio for protein score: 2.14 (CI: 1.54, 2.97).

Of note, hsCRP was not statistically significant in both baseline and baseline + protein score models (p-value>0.05).

Integrated Discriminant Index when protein score was added to the baseline* model was 0.08 (0.05, 0.10).

Table S8: Incremental contribution of protein factors to CVD risk when added to baseline model. Analyses restricted to PLWH on ART at baseline, n=269

Model	AUC	LCI	UCI	Change in AUC from Baseline	NRI among cases	NRI among controls	Overall NRI
Baseline*	0.70	0.60	0.80	-	-	-	-
+ FAM3B	0.69	0.59	0.79	-0.01	0.09 (-.12,.29)	0.01 (-.14,.15)	0.09 (-.16,.34)
+ ITGA11	0.69	0.59	0.79	-0.01	0.13 (-0.07,0.33)	-0.01 (-0.15,0.14)	0.12 (-0.13,0.38)
+ IL6	0.69	0.59	0.79	-0.01	0.35 (0.16,0.54)	0.18 (0.03,0.32)	0.523 (0.28,0.76)
+ HGF	0.69	0.59	0.79	-0.01	0.09 (-0.12,0.29)	0.09 (-0.06,0.23)	0.17 (-0.08,0.42)
+ CCL25	0.70	0.60	0.80	0.01	-0.02 (-0.23,0.18)	0.11 (-0.04,0.25)	0.09 (-0.17,0.34)
+ GT	0.71	0.61	0.81	0.01	0.30 (-0.11,0.50)	0.10 (-0.05,0.24)	0.40 (0.16,0.64)
+ PLA2G7	0.69	0.59	0.79	-0.01	0.17 (-0.03,0.38)	0.13 (-0.02,0.28)	0.30 (0.06,0.55)
+SCGB3A2	0.71	0.61	0.81	0.01	0.15 (-0.05,0.35)	0.11 (-0.04,0.25)	0.26 (0.01,0.51)
+ Protein score from all 8 proteins	0.74	0.64	0.83	0.04	0.28 (-0.09, 0.48)	0.16 (-0.02, 0.31)	0.45 (-0.20, 0.69)

Baseline* is a model with the following variables: CD4, RNA, sex, age, BMI, diabetes status at baseline, prior history of CVD, lipid lowering medication, and blood pressure lowering medication.

The proteins IL6, HGF, and GT were each statistically significant (p-value<0.05) when added to the baseline model. PLA2G7 was marginally significant (p-value =0.06).

Protein score was statistically significant ($p < 0.01$) when added to the baseline model. Odds ratio for protein score: 1.93 (CI: 1.35-2.77); Integrated Discriminant Index: 0.06 (0.04-0.09)

Table S9: Incremental contribution of protein factors to CVD risk when added to baseline model. Analyses restricted to individuals with no prior history of CVD (n=356)

Model	AUC	LCI	UCI	Change in AUC from Baseline	NRI among cases	NRI among controls	Overall NRI
Baseline*	0.66	0.56	0.75	-	-	-	-
+ FAM3B	0.67	0.58	0.76	0.01	0.24 (0.06, 0.42)	0.04 (-0.08, 0.16)	0.28 (0.06,0.50)
+ ITGA11	0.65	0.56	0.75	-0.01	0.05 (-0.13,0.24)	0.00 (-0.12,0.12)	0.05 (-0.17, 0.28)
+ IL6	0.67	0.58	0.77	0.01	0.33 (0.15, 0.50)	0.17 (0.05, 0.29)	0.50 (0.28, 0.71)
+ HGF	0.67	0.58	0.77	0.01	0.22 (0.04, 0.40)	0.11 (-0.01, 0.24)	0.33 (0.11, 0.55)
+ CCL25	0.67	0.57	0.76	0.01	0.11 (-0.08, 0.30)	0.12 (-0.00, 0.25)	0.23 (0.01, 0.45)
+ GT	0.68	0.59	0.77	0.02	0.24 (0.06,0.42)	0.08 (-0.04, 0.21)	0.32 (0.10, 0.54)
+ PLA2G7	0.70	0.61	0.79	0.04	0.22 (0.04, 0.40)	0.11 (-0.02, 0.23)	0.32 (0.10, 0.54)
+SCGB3A2	0.68	0.59	0.78	0.03	0.18 (-0.00, 0.37)	0.18 (0.06, 0.30)	0.36 (0.14, 0.58)
+ Protein score from all 8 proteins	0.70	0.61	0.79	0.04	0.46 (0.29, 0.62)	0.26 (0.14,0.38)	0.72 (0.51, 0.92)

Baseline* is a model with the following variables: CD4, RNA, sex, age, BMI, diabetes status at baseline, lipid lowering medication, and blood pressure lowering medication.

The proteins FAM3B, IL6, HGF, CCL25, GT, PLA2G7 and SCGB3A2 were each statistically significant (p-value<0.05) when added to the baseline model.

Protein score was statistically significant ($p < 0.00001$) when added to the baseline model. Odds ratio for protein score: 2.16 (CI: 1.57-2.96)

Integrated Discriminant Index when protein score was added to baseline model: 0.010 (0.07-0.12)

Table S10: Conditional logistic regression model of CVD on standardized protein score restricted to patients with smoking status reported at baseline. The sample size is 149 (50 CVD cases and 99 CVD controls). Here, we adjust for smoking status at baseline.

Variable	Odds Ratio	Std Error	Pr(>|z|)	LCI	UCI
CD4	1.00	0.00	0.62	1.00	1.00
RNA	1.00	0.00	0.56	1.00	1.00
Females vs Males	0.26	0.78	0.09	0.06	1.21
Age	1.25	0.11	0.04	1.01	1.53
BMI	1.06	0.06	0.34	0.95	1.18
Diabetes	0.63	0.93	0.62	0.10	3.90
Black Race	1.93	0.61	0.28	0.58	6.43
Lipid-lowering medication	1.42	0.71	0.63	0.35	5.71
BP-lowering medication	1.02	0.55	0.97	0.35	2.98
CVD	2.55	0.96	0.33	0.39	16.62
Smoking status (Yes vs No)	2.11	0.62	0.23	0.63	7.10
Protein Score	2.85	0.35	<0.01	1.45	5.63

Table S11: Incremental contribution of protein factors to CVD risk when added to baseline model that includes smoking status. The sample size is 149 (50 CVD cases and 99 CVD controls).

Model	AUC	LCI	UCI	Change in AUC from Baseline	NRI among cases	NRI among controls	Overall NRI
Baseline* (includes smoking status)	0.79	0.68	0.91	-	-	-	-
+ FAM3B	0.84	0.73	0.94	0.04	0.48(0.24,0.72)	0.07(-0.12,0.27)	0.55(0.24,0.86)
+ ITGA11	0.77	0.65	0.90	-0.02	0.24(-0.03,0.51)	0.05(-0.14,0.25)	0.29(-0.04,0.62)
+ IL6	0.77	0.65	0.90	-0.02	0.12(-0.16,0.40)	0.09(-0.10,0.29)	0.21(-0.13,0.55)
+ HGF	0.76	0.64	0.89	-0.03	0.28(0.01,0.55)	0.15(-0.04,0.34)	0.43(0.10,0.76)
+ CCL25	0.81	0.70	0.93	0.02	0.12(-0.16,0.40)	0.13(-0.06,0.32)	0.25(-0.08,0.59)
+ GT	0.81	0.70	0.93	0.02	0.08(-0.20,0.36)	0.11(-0.08,0.30)	0.19(-0.15,0.53)
+ PLA2G7	0.82	0.71	0.94	0.03	0.20(-0.07,0.47)	0.09(-0.10,0.29)	0.29(-0.04,0.62)
+SCGB3A2	0.77	0.65	0.89	-0.02	-0.12(-0.40,0.16)	0.13(-0.06,0.32)	0.01(-0.32,0.35)
+ Protein score from all 8 proteins	0.81	0.69	0.94	0.02	0.44(0.19, 0.69)	0.19(0.00,0.38)	0.63 (0.32, 0.95)

Baseline* is a model with the following variables: CD4, RNA, sex, age, BMI, diabetes status at baseline, prior history of CVD, lipid lowering medication, blood pressure lowering medication, and smoking status.

No individual protein was statistically significant (p-value> 0.05) when added to the baseline model that included smoking status.

Protein score was statistically significant (p =0.0025) when added to the baseline model that included smoking status. Odds ratio for protein score: 2.56(CI: 1.39 – 4.70)

Integrated Discriminant Index when protein score was added to baseline model: 0.10 (0.05-0.14)

Table S12: Odds ratios and AUCs for discovery and validation sets.

Model	Discovery	Internal Validation
Baseline	0.69 (95% CI:0.60-0.78)	0.74 (95% CI: 0.73-0.74)
Baseline + Protein Score	0.73 (95% CI:0.65-0.81)	0.79 (95% CI: 0.78- 0.80)

Odds ratios for protein score in validation set is 2.78 (CI: 2.67, 2.90) and discovery set is 2.17 (CI: 1.58-2.99)
Baseline* is a model with the following variables: CD4, RNA, sex, age, BMI, diabetes status at baseline, prior history of CVD, lipid lowering medication, and blood pressure lowering medication,

Figure S1: Schematic representation of statistical approach

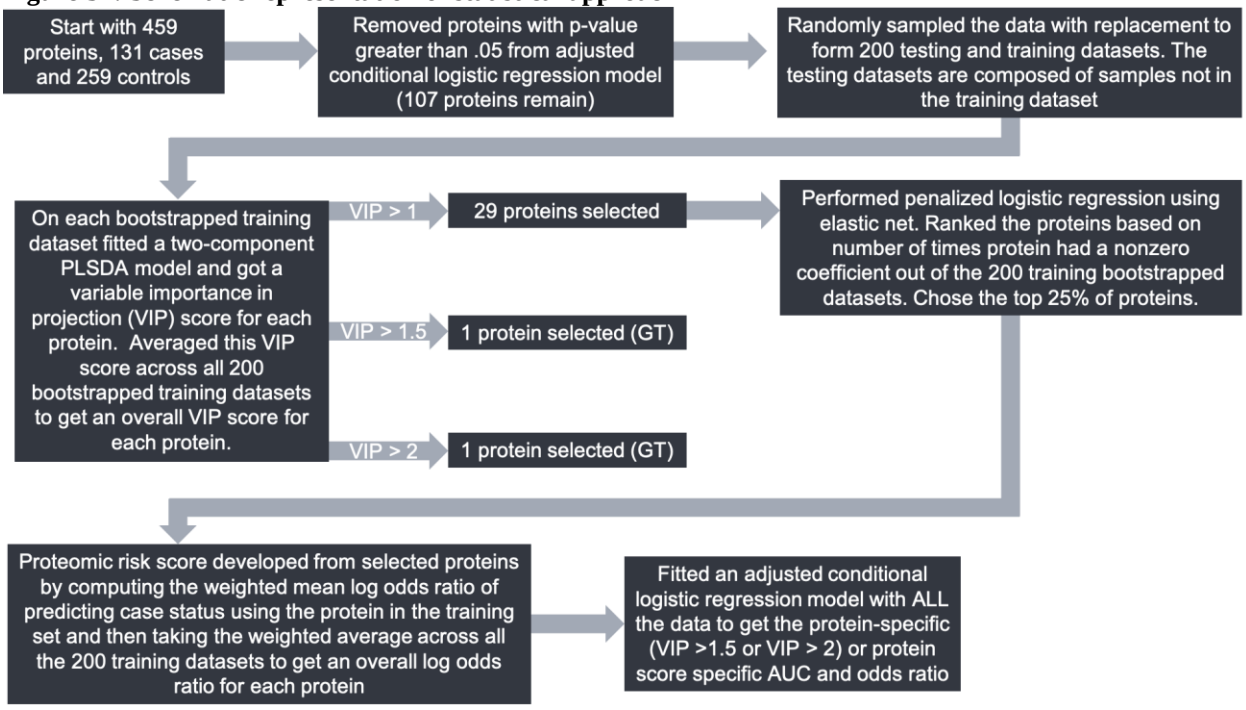
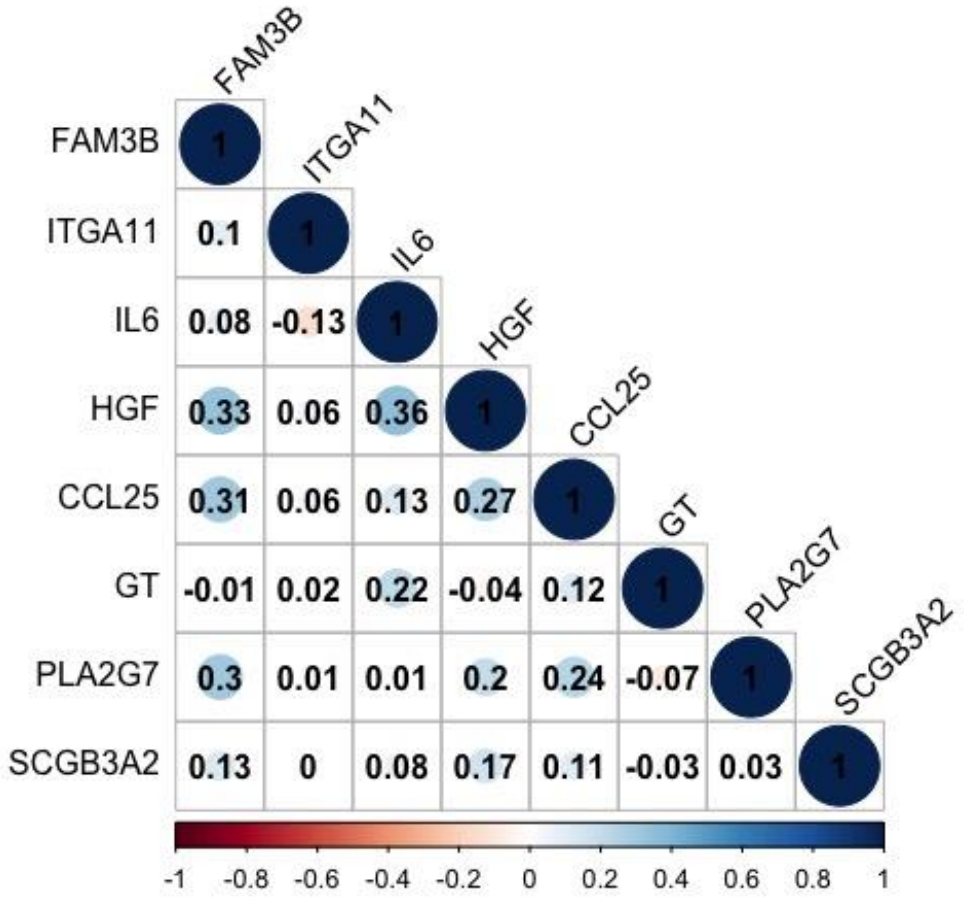


Figure S2: Correlation between pairs of proteins used in the protein score.



C. Description of Proteins Used in the Protein Score

Inflammation Panel:

Hepatocyte Growth Factor (HGF):
The protein HGF and its receptor c-MET are involved in tissue repair and respond to tissue injury. HGF has been proposed as a potential clinical biomarker for CVD[45] HGF has already been shown to be associated with stroke, CHD, atherosclerosis, and the progression of atherosclerosis in an ethnically diverse, general population.[31,32]

C-C motif chemokine 25 (CCL25):
CCL-25 is involved in T-cell development. T-cells are involved in the development and progression of CVDs, including atherosclerosis.[36]

Immune Response Panel:

Protein FAM3B (FAM3B):
Vascular smooth muscle cells (VSMCs) play an important role in the development of CVDs. [35] FAM3B is secreted with insulin and regulates glucose homeostasis. As such, dysregulation of FAM3B is associated with diabetes, which is a risk factor known to be associated with CVD. [35]

Integrin α11 (ITGA11):
Integrins have two functions: 1) extracellular matrix cell attachment and 2) signal transduction from the extracellular matrix.[37] In mouse models, overexpression of integrin α11 induces cardiac fibrosis and left ventricular hypertrophy.[38]

Interleukin-6 (IL6):
IL6 has been extensively studied in both healthy and HIV positive populations. IL6 is a marker of inflammation and coagulation. Increased levels of plasma IL6 has been shown to associated with increased risk of CVD, atherosclerosis, and mortality in an HIV positive population even when treated with ART[17,29,30]

Cardiovascular 2 Panel:

Gastrotropin (GT):
Gastrotropin, also known as the ileal fatty acid binding protein, (FABP6) is a member of the fatty acid-binding protein (FABPs) family, which regulates general metabolic function via FABPs central role in fatty acid transport, metabolism, and storage. FABPs have been associated with a number of diseases including cardiovascular disease and are thought to serve an integral role in metabolic function.[39] FABP6 is more specifically known to be involved in bile acid metabolism. There has been shown to be a protective association between FABP6 Thr79Met polymorphism and incident type 2 diabetes.[40]

Cardiovascular 3 Panel

Platelet-activating factor acetylhydrolase (PLA2G7):

PLA2G7 is found in both high-density lipoprotein (HDL) and low-density lipoprotein (LDL). In population studies it has been shown that overexpression of PLA2G7 is associated with increased coronary heart disease (CHD). [33, 34] It is thought that with individuals with low LDL cholesterol levels it can help predict CHD risk. [33,34]

Cardiometabolic Panel

Secretoglobin Family 3A Member (SCGB3A2):

A case-control study in a Korean population on SNPs in the SCGB3A2 gene potentially contributes to susceptibility to asthma. [41] The SCGB3A2 gene is also known as the uterus globulin associated protein 1 (UGRP1) found that UGRP1 may be able to predict graves' disease patients who develop hypothyroidism.[42]

D. **Sensitivity and Subgroup Analyses**

When we added the inflammatory biomarkers D-dimer and hsCRP individually (as $\log_2$-transformed measures) to the baseline model and we considered the protein score + baseline model, the protein score was again statistically significant ($p < 0.0001$) [Tables S6-S7]. Further, the AUCs and NRIs estimates from the baseline + protein score model were again higher than those from the baseline + individual proteins models. With the exception of the protein SCGB3A2, the proteins FAM3B, IL-6, HGF, GT, and PLA2G7 were each statistically significant (p-value<0.05) when added to the baseline model.

Considering the 269 participants with complete data (92 cases and 177 controls) who were on ART at baseline, we observed that a one standard deviation increase in the protein score was associated with an odds ratio for CVD of 1.93 (CI: 1.35 - 2.78) [Table S8] compared to 2.17 (CI: 1.58 - 2.99) [Figure 1] in the combined data (i.e. data on those with ART and no ART at baseline). Unlike in the combined data where IL6 was likely driving the effect of the protein score, in this subpopulation, no individual protein had an AUC that was comparable to the protein score (Table S8). Further, the AUC for the baseline model was 0.70 (CI: 0.60- 0.80), and the AUC for the baseline + protein score model was 0.74(CI: 0.64 - 0.83). There was a 16.7% increase in prediction when the analysis was restricted to those on ART at baseline, compared to a 17.4% increase in prediction for the full population. The NRI for the protein score + baseline model was 0.45 (CI: 0.20 - 0.69) in this subgroup as compared to 0.66 (CI: 0.46 - 0.86) in the combined data. Table S8 gives the NRI and IDI estimate for the protein score + baseline model restricted to individuals on ART at baseline as well as the incremental contributions of each protein.

When we restricted our analyses to individuals with no history of CVD (n=356, 110 cases and 246 controls), the protein score was again statistically significant (Table S9). A one standard deviation increase in the protein score was associated with an odds ratio for CVD of 2.16 (CI: 1.57-2.96) compared to 2.17 (CI: 1.58 - 2.99) in the combined data. The AUC for the baseline model was 0.66 (CI: 0.56-0.75). The AUC when PLA2G7 was added to the baseline model was the same as the AUC when the protein score was added to the baseline model (0.70) but the protein score + baseline model had a higher NRI (0.72 vs 0.32); the NRI for the full population was 0.66 (CI: 0.46 - 0.86).

We considered a sensitivity analysis that included individuals with smoking status. Here, the sample size was 149 with 50 CVD cases and 99 CVD controls. When we added smoking status to the baseline model, the protein score from the protein score + baseline model was statistically significant (p-value =0.0024) [Table S10]. On the other hand, no individual protein achieved statistical significance when each was added to the baseline model that included smoking status as a covariate. A one standard deviation increase in the protein score was associated with an odds ratio for CVD of 2.85 (CI: 1.45- 5.63). The AUC for the baseline model was 0.79 (CI: 0.68- 0.91), and the AUC for the baseline + protein score model was 0.81 (CI: 0.69 - 0.94). This corresponded to a 6.8% increase in prediction compared to a 17.4% increase for the full population. Further, FAM3B, CCL25, GT, and PLA2G7 had AUC values that were similar or slightly higher than the AUC from the protein score + baseline model (Table S11). However, the NRI for these individual proteins were lower than the NRI from the baseline + protein score model.