Check for updates

# Plasma proteomics and carotid intima-media thickness in the UK biobank cohort

Ming-Li Chen[1], Pik Fang Kho[1], Rodrigo Guarischi-Sousa[1,2], Jiayan Zhou[1], Daniel J. Panyard[3], Zahra Azizi[1], Trisha Gupte[1], Kathleen Watson[4], Fahim Abbasi[1,5] and Themistocles L. Assimes[1,5]*

[1]Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA, United States, [2]Palo Alto Veterans Institute for Research (PAVIR), Stanford, CA, United States, [3]Department of Genetics, Stanford University School of Medicine, Stanford, CA, United States, [4]Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, United States, [5]Stanford Cardiovascular Institute, Stanford University School of Medicine, Stanford, CA, United States

**Background and aims:** Ultrasound derived carotid intima-media thickness (cIMT) is valuable for cardiovascular risk stratification. We assessed the relative importance of traditional atherosclerosis risk factors and plasma proteins in predicting cIMT measured nearly a decade later.

**Method:** We examined 6,136 UK Biobank participants with 1,461 proteins profiled using the proximity extension assay applied to their baseline blood draw who subsequently underwent a cIMT measurement. We implemented linear regression, stepwise Akaike Information Criterion-based, and the least absolute shrinkage and selection operator (LASSO) models to identify potential proteomic as well as non-proteomic predictors. We evaluated our model performance using the proportion variance explained ($R^2$).

**Result:** The mean time from baseline assessment to cIMT measurement was 9.2 years. Age, blood pressure, and anthropometric related variables were the strongest predictors of cIMT with fat-free mass index of the truncal region being the strongest predictor among adiposity measurements. A LASSO model incorporating variables including age, assessment center, genetic risk factors, smoking, blood pressure, trunk fat-free mass index, apolipoprotein B, and Townsend deprivation index combined with 97 proteins achieved the highest $R^2$ (0.308, 95% C.I. 0.274, 0.341). In contrast, models built with proteins alone or non-proteomic variables alone explained a notably lower $R^2$ (0.261, 0.228−0.294 and 0.260, 0.226−0.293, respectively). Chromogranin b (CHGB), Cystatin-M/E (CST6), leptin (LEP), and prolargin (PRELP) were the proteins consistently selected across all models.

**Conclusion:** Plasma proteins add to the clinical and genetic risk factors in predicting a cIMT measurement. Our findings implicate blood pressure and extracellular matrix-related proteins in cIMT pathophysiology.

## Introduction

Ultrasound derived carotid intima media thickness (cIMT) is well-established tool used to noninvasively measure the thickness of the carotid artery in high-resolution (1). Suggested applications of cIMT include using the measurement to improve cardiovascular risk stratification and the prediction of future cardiovascular events over

that possible with traditional risk factors alone, as well as the assessment of the efficacy of cardiovascular drugs (2–5). Nevertheless, the reproducibility of cIMT images has long been its major limitation because performing cIMT measurements and image analysis requires standardized equipment and protocol, as well as a trained sonographer (6).

High throughput profiling of circulating proteins in the plasma has emerged as a useful approach to improved risk prediction as well as to provide novel insights on the biology of complex human traits (7). When integrated with genetic and environmental determinants of diseases, proteomic-based models can outperform clinical risk factor models in predicting cardiovascular events (8, 9). However, it is unknown to what extent plasma proteomics correlate with cIMT, and whether they improve its prediction over clinical measures.

We aimed to evaluate the potential of a signature of plasma proteins to predict the measurement of cIMT nearly a decade later in a generally healthy population. To accomplish this task, we leveraged data from the UK Biobank (UKBB) to investigate individuals with measurements of circulating proteins in the plasma, genetic variation, lifestyle factors, and health outcomes, and ultrasound imaging of the carotids (10).

# Methods

## Study population

We analyzed a sub cohort of UKBB participants of the UKBB who underwent both high-throughput proteomic profiling as well as a cIMT study (Figure 1A). The study design, as well as the methods of data collection of the UKBB is described elsewhere in detail (11, 12). Briefly, participants' baseline visit occurred between 2006 and 2010 (Instance 0) during which baseline characteristics and a personal health history were collected using touchscreen questionnaires, physical measures, and verbal interviews. Blood was also drawn and banked. A subset of participants returned for an imaging visit starting in 2014 (Instance 2). The UKBB received ethical approval from the National Information Governance Board for Health and Social Care and the Northwest Multicenter Research Ethics Committee. All participants provided their written informed consent at baseline.

## Plasma proteomic biomarkers

A total of 1,463 circulating proteins in plasma were previously measured using the antibody-based Proximity Extension Assay by Olink. Extensive quality control of protein measures was implemented by the UKBB using a set of predefined criteria as previously described (10). Levels were normalized using a two-step approach of within-batch and across-batches intensity normalization (13). We excluded data for two proteins with a high missing rate across our study samples (NPML with 74% and PCOLCE with 63% missing). The values were log-transformed and standardized before analysis. Missing normalized protein expression (NPX) values were imputed with their mean values.

## Outcome

The cIMT was measured at the imaging visit (instance 2) in four angles using the CardioHealth Station ultrasound system and UKBB predefined standards (14, 15). We included only individuals with cIMT measurements that did not fail any quality control indicators in our analyses. The average value of cIMT of the four angles was taken as the primary outcome measure in this study. The distribution of cIMT measures was acceptable (skewness = 0.89) after removing outliers of larger than five standard deviations (SD) from the mean. A detailed description of how the cIMT measurement was acquired is included in the Supplementary File.

## Selection of non-proteomic covariates

Potential non-proteomic covariates were selected based on a careful literature review focused on the identification of previously established and/or suspected non-proteomic predictors of cIMT. We considered the following groups of variables documented at baseline: demographic characteristics (sex, age when attending the baseline assessment and the imaging assessment, self-reported race/ethnicity, Townsend deprivation index, and assessment center location), anthropometric characteristics (waist circumference, body mass index, calculated whole body and trunk fat as well as fat-free mass index estimated through whole-body bio-impedance), health behaviors (alcohol consumption, ever smoked, and categorized physical activity groups), medical conditions, medications, physical measures (blood), nine biochemistry markers, five genetic principle components, and four polygenic risk scores (PRS) of health traits related to cardiometabolic risk from the "Standard PRS Set" UKBB resource (16, 17).

## Statistical analyses

We first determined the marginal association between each non-proteomic factor and cIMT. Subsequently, we selected five non-modifiable characteristics (sex, age including at baseline and at the imaging visit, center of imaging visit, and self-reported ethnicity) to define a minimally adjusted model. We then applied an Akaike information criterion (AIC) based stepwise model selection algorithm to the remaining non-proteomic covariates of interest to identify a subset that improved the AIC for cIMT in a multivariable variable framework. Non-proteomic variables that were selected were then further grouped into four to allow for multivariate testing of five linear regression models (LM) with increasingly progressive covariate adjustment: (1) the five basic non-modifiable characteristics (LM 1), (2) model 1 plus multiple PRS (LM 2), (3) model 2 plus health behaviors (LM 3), (4) model 3 plus physical measures (LM 4), and (5) model 4 plus medical conditions (LM 5). More detailed analysis plan is described in Supplementary File.
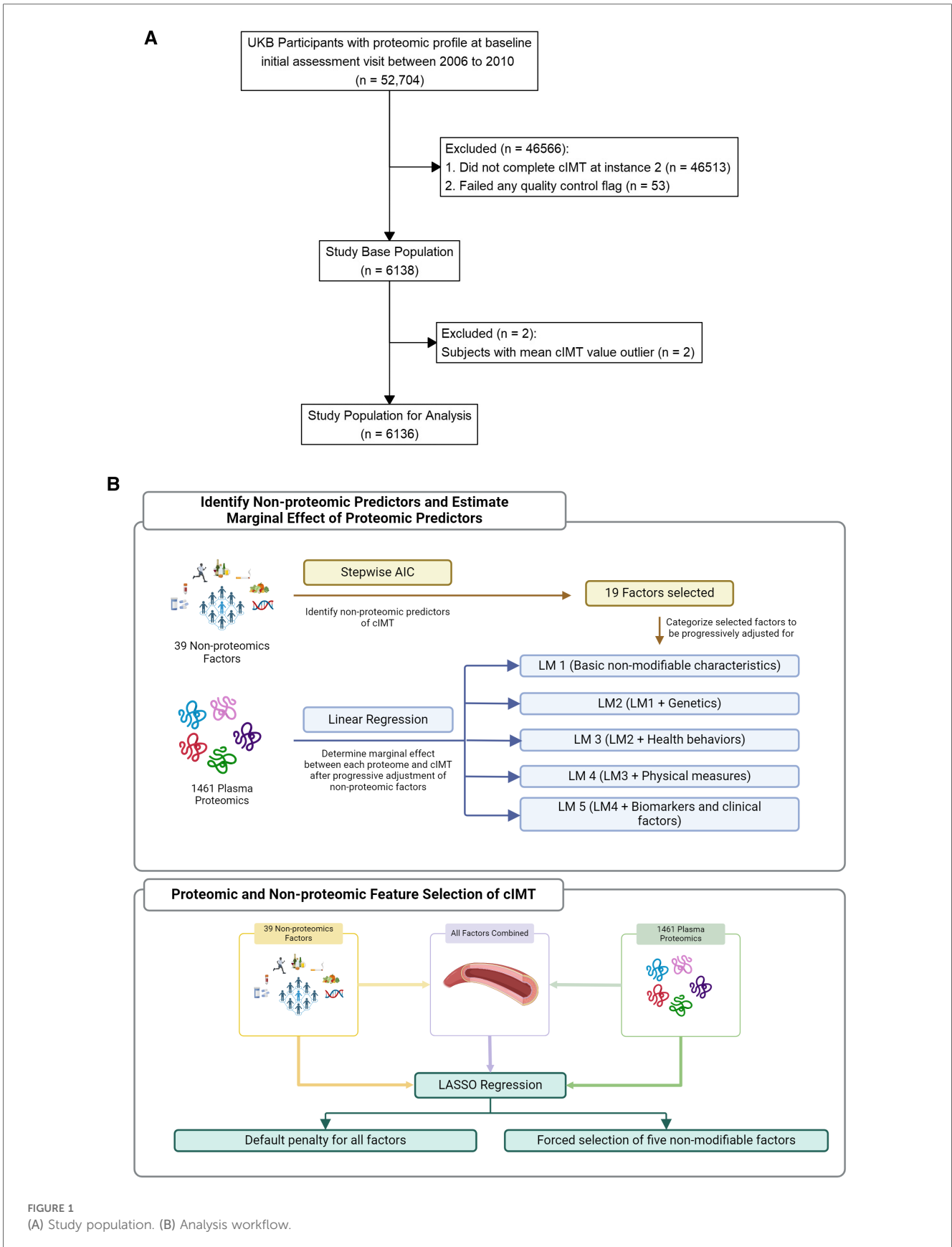
**FIGURE 1**
(A) Study population. (B) Analysis workflow.

Next, we used the least absolute shrinkage and selection operator (LASSO) regression to identify a subset of plasma proteomic measures predictive of cIMT alone, or in combination with non-proteomic covariates. We built the LASSO model using 10-fold cross-validation after creating a randomly selected training (70%) and testing (30%) subset. Factors were centered and scaled before analysis. We trained six LASSO models, with a pair of models drawing only from protein levels, a pair drawing only from non-proteomic variables, and a pair drawing from a combination of both. Within each pair, one model was trained after forcing the five non-modifiable characteristics of model 1 without penalty. Lastly, we quantified model improvement after adding proteomic measures to other non-proteomic variables by evaluating improvements in variance explained ($R^2$) and reductions in the root mean square error (RMSE) in the testing set. The 95% confidence intervals were generated by bootstrapping 1,000 samples. The Benjamini-Hochberg False Discovery Rate (FDR) threshold of 0.05 was used to account for multiple testing of proteomic-cIMT associations.

Sensitivity analyses were performed excluding protein measurements with a high proportion of NPX values flagged as being below the limit of detection (LOD) corresponding to each sample's plate ID. We repeated the six LASSO models in the same training and testing population after excluding proteins with below LOD measures for >25%, >10%, and >3% of the participants, respectively. 95% confidence intervals were generated by bootstrapping 1,000 samples. Statistical analyses were performed with R, version 4.1.1.

## Gene set enrichment analysis

We implemented gene set enrichment analysis (GSEA) to investigate the possible enrichment of biochemical pathways from proteins identified in the previous screening method (18–22). For this analysis, the normalized enrichment score (NES) was the primary statistic. We considered two levels of significance, an FDR q value of less than 0.05 and a more liberal FDR q value of less than 0.25 if no gene sets met the more stringent threshold of significance. GSEA software (v4.3.2) was used for this analysis.

## Results

### Cohort characteristics

A total of 1,461 plasma proteins levels in 6,136 participants were included in our analysis (Figure 1A). The baseline characteristics of the study population are shown in Table 1. The mean time from baseline visit to imaging visit for cIMT visit was approximately 9 years with a little less than half of the participants being men. Summary statistics of the full list of covariates considered are included in Supplementary Table 1 while additional annotation of proteins measured including their summary statistics within the study cohort is included in Supplementary Table 2.

TABLE 1 Demographic and clinical characteristics of total study population.

| | Overall |
|---|---|
| $n$ | 6,136 |
| Male (%) | 2,901 (47.3) |
| Age at baseline | 54.39 (7.86) |
| Age at imaging visit | 63.61 (8.08) |
| Ethnicity group (%) | |
| Asian | 67 (1.1) |
| Black | 51 (0.8) |
| Mixed | 41 (0.7) |
| Other | 52 (0.8) |
| White | 5,925 (96.6) |
| Townsend deprivation index at recruitment [median (IQR)] | −2.56 [−3.88, −0.24] |
| Ever smoked (%) | 3,597 (58.6) |
| Alcohol intake frequency (%) | |
| Never or missing | 296 (4.8) |
| One to three times a month or special occasions only | 1,154 (18.8) |
| Once or twice a week | 1,564 (25.5) |
| Three or four times a week | 1,727 (28.1) |
| Daily or almost daily | 1,395 (22.7) |
| Physical activity category (%) | |
| Missing | 884 (14.4) |
| Low | 981 (16.0) |
| Moderate | 3,648 (59.5) |
| High | 1,507 (24.6) |
| Waist circumference (cm) | 87.61 (12.43) |
| BMI (kg/m$^2$) | 26.48 (4.21) |
| Whole body fat mass index (kg/m$^2$) | 8.10 (3.19) |
| Whole body fat-free mass index (kg/m$^2$) | 18.40 (2.51) |
| Trunk fat mass index (kg/m$^2$) | 4.47 (1.64) |
| Trunk fat-free mass index (kg/m$^2$) | 10.25 (1.26) |
| SBP (mmHg) | 135.96 (18.54) |
| DBP (mmHg) | 80.99 (10.55) |
| LDL direct [mmol/L, median (IQR)] | 3.50 [2.96, 4.09] |
| HDL cholesterol [mmol/L, median (IQR)] | 1.43 [1.21, 1.71] |
| TG [mmol/L, median (IQR)] | 1.38 [0.98, 1.98] |
| Cholesterol [mmol/L, median (IQR)] | 5.63 [4.94, 6.39] |
| HbA1c [mmol/mol, median (IQR)] | 34.50 [32.10, 37.00] |
| Apolipoprotein A [g/L, median (IQR)] | 1.52 [1.36, 1.70] |
| Apolipoprotein B (g/L, median [IQR] | 1.00 [0.86, 1.17] |
| Lipoprotein A [nmol/L, median (IQR)] | 19.96 [9.34, 57.83] |
| C-reactive protein [mg/L, median (IQR)] | 1.05 [0.54, 2.10] |
| Vascular/heart problems diagnosed by doctor (%) | 1,278 (20.8) |
| Diabetes diagnosed by doctor (%) | 137 (2.2) |
| Blood pressure medication (%) | 497 (8.1) |
| Cholesterol lowering medication (%) | 492 (8.0) |
| Statin use (%) | 628 (10.2) |
| Aspirin use (%) | 661 (10.8) |
| PRS for CVD | −0.15 (0.95) |
| PRS for HTN | −0.11 (0.93) |
| PRS for ISS | −0.08 (0.91) |
| PRS for T2D | −0.22 (0.95) |
| Follow up time (days) | 3,369.91 (697.15) |
| Mean cIMT (micrometers) | 680.67 (123.99) |

All continuous measurements were documented in mean (SD) unless otherwise specified. IQR, interquartile range; BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; TG, triglycerides; HbA1c, glycated hemoglobin; PRS, polygenic risk score; CVD, cardiovascular disease; HTN, hypertension; ISS, ischemic stroke; T2D, type 2 diabetes; PC, principal components.

## Covariate selection with AIC and standard linear regression of proteins

From a total of 39 non-proteomic factors analyzed, the most predictive covariates in univariate regression models were those related to age (either at baseline or at the time of imaging visit) with each explaining nearly one fifth of the variance ($R^2$) of cIMT (Supplementary Table 3). In contrast, the next strongest non-modifiable predictor was sex which explained only between 2% and 3% of $R^2$. The most predictive modifiable covariate was systolic blood pressure explaining about 6%–7% of $R^2$. A total of 19 out of 39 potential covariates were selected using the AIC based stepwise model selection algorithm and grouped into the five LM (Supplementary Table 4). To benchmark the predictive ability of these non-proteomic variables selected in the previous five multiple LMs, we calculated the variance explained ($R^2$) of each model in the testing dataset. As more explanatory demographic, genetic, and phenotypic variables were included, especially blood pressure and anthropometric measures, the $R^2$ increased from 0.211 (95%C.I. 0.178, 0.244) in LM 1 to 0.256 (95%C.I. 0.223, 0.290) in the fully adjusted LM 5 (Supplementary Figure 1).

Next, we screened all proteins to identify the subset of proteins marginally associated with cIMT with progressive adjustment of our non-proteomic covariates identified. When LM 1 through 5 were applied in multivariable model with each protein, we identified 374, 289, 310, 11, and 4 proteins at FDR < 0.05. The 11 proteins in LM 4 that were significant included Secretogranin-1 (CHGB), Cystatin-M (CST6), formin-like protein 1 (FMNL1), interleukin-17D (IL17D), leptin (LEP), matrix extracellular phosphoglycoprotein (MEPE), phosphoprotein associated with glycosphingolipid-enriched microdomains 1 (PAG1), platelet-activating factor acetylhydrolase (PLA2G7), phospholipid transfer 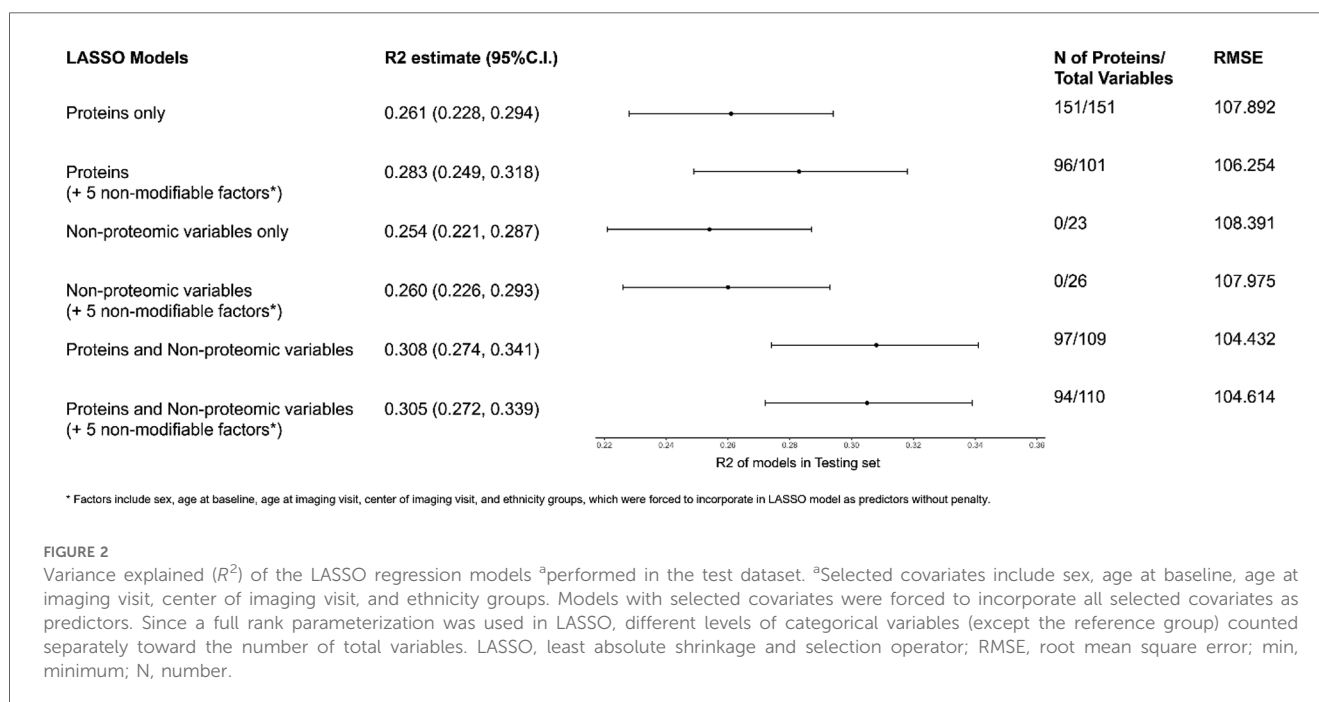protein (PLTP), prolargin (PRELP), and signaling threshold-regulating transmembrane adapter 1 (SIT1) while the four proteins that remained significant in LM 5 where CHGB, CST6, LEP, and PRELP (Supplementary Table 5).

## LASSO regression models

Using LASSO regression, we noted the following trends (Figure 2; Supplementary Table 6). In models that allowed for the selection of proteins, a range of 94–151 proteins were selected while in models with non-proteomic covariates, a range of 12–16 non-proteomic covariates were selected. A model restricted to non-proteomic measures explained the least variance ($R^2 = 0.254$). A model including the five non-modifiable covariates in addition to proteins increased the variance explained ($R^2 = 0.283$) to a degree that was comparable to models built from a combination of non-proteomic and proteomic predictors. A LASSO model combining 97 proteins with 12 non-proteomic factors performed the best in the testing set ($R^2 = 0.308$). Lastly, we observed significant overlap of the 95% confidence intervals of all models that included at least some non-proteomic covariates. The cross-validation mean squared error plots of all six LASSO models for the study population are shown in Supplementary Figure 2. A summary of all proteomic and non-proteomic predictors selected by the linear regression or LASSO models is shown in Supplementary Tables 7, 8.

## Sensitivity analysis LASSO regression models

A total of 175 out of 1,461 proteins had more than 25% of their measures below LOD, while 244 proteins had more than 10% and 472 proteins had more than 3% below LOD. When excluding these



| LASSO Models | R2 estimate (95%C.I.) | | N of Proteins/ Total Variables | RMSE |
|---|---|---|---|---|
| Proteins only | 0.261 (0.228, 0.294) | | 151/151 | 107.892 |
| Proteins (+ 5 non-modifiable factors*) | 0.283 (0.249, 0.318) | | 96/101 | 106.254 |
| Non-proteomic variables only | 0.254 (0.221, 0.287) | | 0/23 | 108.391 |
| Non-proteomic variables (+ 5 non-modifiable factors*) | 0.260 (0.226, 0.293) | | 0/26 | 107.975 |
| Proteins and Non-proteomic variables | 0.308 (0.274, 0.341) | | 97/109 | 104.432 |
| Proteins and Non-proteomic variables (+ 5 non-modifiable factors*) | 0.305 (0.272, 0.339) | | 94/110 | 104.614 |

R2 of models in Testing set

\* Factors include sex, age at baseline, age at imaging visit, center of imaging visit, and ethnicity groups, which were forced to incorporate in LASSO model as predictors without penalty.

FIGURE 2
Variance explained ($R^2$) of the LASSO regression models [a]performed in the test dataset. [a]Selected covariates include sex, age at baseline, age at imaging visit, center of imaging visit, and ethnicity groups. Models with selected covariates were forced to incorporate all selected covariates as predictors. Since a full rank parameterization was used in LASSO, different levels of categorical variables (except the reference group) counted separately toward the number of total variables. LASSO, least absolute shrinkage and selection operator; RMSE, root mean square error; min, minimum; N, number.

proteins, the variance explained was largely unchanged compared to that observed in the main analyses (Supplementary Table 9).

## Gene set enrichment analysis

In GSEA, while no gene set was significant at FDR-q <0.05, we found 9 upregulated canonical pathways, 3 downregulated canonical pathways, and 3 Hallmark gene sets reached FDR <0.25 (Supplementary Table 10 and Supplementary Figures 3, 4).

# Discussion

We leveraged the large-scale profiling of nearly 1,500 proteins in the plasma of >6,000 participants of the UK Biobank study who also underwent ultrasound imaging of their carotids to provide an initial estimate of the relative importance of proteomic and non-proteomic measures in predicting a cIMT measurement and in understanding the pathophysiologic link behind an increased rate of thickening and the future risk of atherosclerosis related complications in both proximal and more distant arterial trees (23, 24).

With respect to prediction of cIMT, we document several interesting trends. First, as has been shown by others in the same cohort (25), age was by far the single strongest predictor of cIMT explaining three times more $R^2$ than the next best non-proteomic predictor, systolic blood pressure, and about two thirds the maximal $R^2$ explained by our best LASSO model. Second, approximately one quarter of all proteins measured were marginally associated with cIMT in the setting of our minimally adjusted model. Third, a model restricted to approximately 150 proteins performed as well as the best predictive model using clinical variables explaining a little over a quarter of the $R^2$ of cIMT. Lastly, the addition of approximately 100 proteins through a LASSO model notably increased the absolute $R^2$ explained from ~25% to ~30%, or a ~20% relative increase. While the added predictive value of proteomic profiling was not dramatic, our findings open the door to the future development of a simple, blood-based, proteomic or multi-omics signature that may either enhance or serve as a good proxy to cardiovascular risk assessment with a cIMT measurement (26, 27).

We highlight several modifiable factors that were significant in our multivariate linear regression models and were also consistently selected by LASSO including blood pressure, smoking, apolipoprotein B, and trunk fat-free mass index. Despite the very high correlation of the same mass indices between different body parts, the fat-free mass indices were repeatedly selected over fat mass indices by several models. These findings confirm prior studies demonstrating the importance of traditional risk factors of atherosclerosis in determining cIMT (28–30), as well as more recent studies exploring the relative contributions of correlated measures of body composition (29, 31–33). A robust signature of modifiable risk factors within circulating plasma proteins is also evident in our progressive multivariate linear regression models given the number of

marginally associated proteins dropped dramatically with the progressive addition of more non-proteomic measures.

Four proteins remained significantly associated with cIMT after a comprehensive adjustment with non-proteomic variables that included non-modifiable factors, PRS, health behaviors, physical measures, biomarkers and medical history. The first was CHGB or Chromogranin B (also known as Secretogranin-1) (34, 35). CHGB is a member of the granin family of neuroendocrine secretory proteins, commonly found in the secretory vesicles of neurons and endocrine cells and plays a significant role in the regulation of neuroendocrine secretions (34, 35). While a related protein, Chromogranin A, serves as a precursor to Catestatin, a modulator of the neuroendocrine system that has antihypertensive properties through the inhibition of the release of catecholamines, such a relationship does not exist between Chromogranin B and blood pressure (34, 35).

CST6 (also known as Cystatin M/E or Cystatin E/M), is a secreted protein that is part of the type 2 cystatin subfamily. Its primary role is to act as an inhibitor of lysosomal cysteine proteases including cathepsins and legumain by forming high-affinity reversible complexes (36, 37). CST6 is known to contribute to the homeostatis of the skin, but also been implicated in cancer biology, as it is often found to be downregulated, and sometimes completely silenced, in various cancer types such as breast cancer, melanoma, and lung cancer (36, 37). When CST6 isn't functioning properly, it can lead to changes in the proteolysis of tissue structures, which could potentially speed up the spread of cancer cells. These extracellular effects on the tumor microenvironment may also be reflected some effects on the extracellular matrix within the carotid media. Of note, a prior study found Cystatin N/E protein expressions to be positively correlated with symptoms among patients with carotid artery disease (38).

The third protein, the adipocyte derived hormone leptin (LEP), has more obvious connections to cIMT through established effects on blood pressure (39). These effects are facilitated by the enhancement of the sympathetic nervous system, initiated by the stimulation of pro-opiomelanocortin (POMC) neurons in the hypothalamic arcuate nucleus. These neurons project to the periventricular nucleus, leading to the release of the α-melanocortin stimulating hormone, which subsequently activates melanocortin receptors on presympathetic neurons. Separately, leptin causes sodium retention through a direct action on the renal tubules as well as the renin–angiotensin–aldosterone system via stimulation of aldosterone release in the adrenal cortex (39).

The final protein, PRELP, is a secreted protein found in the aortic extracellular environment (ECM)associated with pre-atherosclerotic lesions (40, 41). PRELP also binds the basement membrane protein heparan sulfate proteoglycan perlecan (HSPG2), which regulates lipid deposition (42). In our analysis, HSPG2 was also shown to be positively associated with cIMT. Lastly, the highest bulk tissue gene expression for PRELP is found in vascular tissues (aorta, coronary artery, and tibial artery) and principle single cell sources of this protein include the fibroblast, endothelial cells, pericytes, and smooth muscle cells (43, 44). Together, these findings suggest possible

contributory role of PRELP in maintaining the structural integrity and remodeling of the ECM within the cardiovascular system including the carotid-intima media region.

We compared our findings to a recently published association analysis between cIMT measurements and the levels of 1,095 proteins measured in blood using the SOMAscan Platform in a subset of 893 participants of the KORA F4 population-based cohort study (45). The researchers found four proteins to be associated with cIMT in a basic age-sex adjusted model, cytoplasmic protein NCK1 (NCK1), insulin-like growth factor-binding protein 2 (IGFBP2), growth hormone receptor (GHR), and GDNF family receptor alpha-1 (GFRA1), with only NCK1 remaining significant in their fully adjusted model (45). In our study, neither NCK1 and GHR were measured using Olink, but both GFRA1 and IGFB2 were significant in our models 1, 2, and 3. A look-up of our top findings in KORA F4 revealed results for only two out of our four top proteins, LEP and CST6 with LEP showing nominally significance in both models (45). This degree of replication between studies is not unexpected given differences in design (cross-sectional vs. longitudinal), power (893 vs. 6,183 subjects), analytic approach to declaring significance (Bonferroni vs. FDR) and measurement platform used (SOMA scan vs. OLINK). We note that multiple studies have shown a bimodal distribution of correlations between the two platforms for proteins measured in both platforms (46–49). While the correlations for LEP have been on the high end, the correlations for CST6 have been on the low end.

The main strengths of our study are the standardized protocols for cIMT, the large sample size, and the comprehensive proteomic profiling using a high-throughput platform known to maintain a high specificity. Limitations are several as well. First, the long time between the time of the proteomic and the cIMT measurements may have limited our ability to correlate more strongly with cIMT. Nevertheless, the significant proteins identified in our study may serve as candidate biomarkers for early prediction of cIMT, conditional on age and other cardiovascular risk factors. Secondly, since no atherosclerotic plaque was measured in the UKBB, the association between proteins and plaque will require additional studies to confirm or refute their linkage to atherosclerosis related health traits. Thirdly, the healthy volunteer bias and the predominantly white population (12) may limit the generalizability of our findings and mandate future studies that are both more ancestrally and socioeconomically diverse. Lastly, we acknowledge two significant challenges to conducting our GSEA. First, many gene sets were only partially represented by protein measures in plasma given only ~1,500 protein measures were available. Second, additional substantial additional statistical noise is likely introduced by moving out of cell and representing genes through their plasma levels of proteins derived from potentially all tissues in the body rather than their expression levels intracellularly or by genetic variation affecting their expression levels. Thus, we suspect our analyses are substantially underpowered. Nevertheless, we provide some insights on potentially relevant biology represented by our highest ranked genes sets in the supplementary appendix.

In summary, we found proteomic profiles in plasma achieve notable incremental prediction over non-proteomic factors of cIMT. This incremental prediction is achieved despite the separation of the two sets of measures (plasma proteins and cIMT) by nearly a decade. Our findings suggest proteomic signatures, possibly combined with other -omic measurements in blood, may one day enhance or serve as a reliable cross-sectional surrogate of a direct measurement of cIMT. Among our most robustly associated proteins, we note proteins involved in the modification of core mechanisms determining cIMT such as blood pressure, but also evidence of the role of multiple cell types active in the vascular wall. Future studies are needed to provide a clearer understanding of these associations, and the relevant mechanisms involved.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ukbiobank.ac.uk.

## Ethics statement

The UKBB received ethical approval from the National Information Governance Board for Health and Social Care and the Northwest Multicenter Research Ethics Committee. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

M-LC: Writing – review & editing, Writing – original draft, Formal Analysis, Conceptualization. PK: Writing – review & editing, Methodology. RG-S: Writing – review & editing, Formal Analysis. JZ: Writing – review & editing, Methodology. DP: Writing – review & editing, Methodology. ZA: Writing – review & editing. TG: Writing – review & editing. KW: Writing – review & editing. FA: Writing – review & editing. TA: Writing – review & editing, Supervision, Investigation.

## Funding

## Acknowledgments

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcvm.2024.1478600/full#supplementary-material

## References

1. de Groot E, van Leuven SI, Duivenvoorden R, Meuwese MC, Akdim F, Bots ML, et al. Measurement of carotid intima-media thickness to assess progression and regression of atherosclerosis. *Nat Clin Pract Cardiovasc Med.* (2008) 5(5):280–8. doi: 10.1038/ncpcardio1163

2. Lorenz MW, Polak JF, Kavousi M, Mathiesen EB, Völzke H, Tuomainen TP, et al. Carotid intima-media thickness progression to predict cardiovascular events in the general population (the PROG-IMT collaborative project): a meta-analysis of individual participant data. *Lancet.* (2012) 379(9831):2053–62. doi: 10.1016/s0140-6736(12)60441-3

3. Den Ruijter HM, Peters SA, Anderson TJ, Britton AR, Dekker JM, Eijkemans MJ, et al. Common carotid intima-media thickness measurements in cardiovascular risk prediction: a meta-analysis. *JAMA.* (2012) 308(8):796–803. doi: 10.1001/jama.2012.9630

4. de Groot E, Jukema JW, Montauban van Swijndregt AD, Zwinderman AH, Ackerstaff RG, van der Steen AF, et al. B-mode ultrasound assessment of pravastatin treatment effect on carotid and femoral artery walls and its correlations with coronary arteriographic findings: a report of the regression growth evaluation statin study (REGRESS). *J Am Coll Cardiol.* (1998) 31(7):1561–7. doi: 10.1016/s0735-1097(98)00170-3

5. Willeit P, Tschiderer L, Allara E, Reuber K, Seekircher L, Gao L, et al. Carotid intima-media thickness progression as surrogate marker for cardiovascular risk: meta-analysis of 119 clinical trials involving 100 667 patients. *Circulation.* (2020) 142(7):621–42. doi: 10.1161/circulationaha.120.046361

6. Martínez JMP, Santos JMG, Martinez MLP, Pastor AM. Carotid intima-media thickness and hemodynamic parameters: reproducibility of manual measurements with Doppler ultrasound. *Med Ultrason.* (2015) 17(2):167–74. doi: 10.11152/mu.2013.2066.172.ci-m

7. Joshi A, Rienks M, Theofilatos K, Mayr M. Systems biology in cardiovascular disease: a multiomics approach. *Nat Rev Cardiol.* (2021) 18(5):313–30. doi: 10.1038/s41569-020-00477-1

8. Hoogeveen RM, Pereira JPB, Nurmohamed NS, Zampoleri V, Bom MJ, Baragetti A, et al. Improved cardiovascular risk prediction using targeted plasma proteomics in primary prevention. *Eur Heart J.* (2020) 41(41):3998–4007. doi: 10.1093/eurheartj/ehaa648

9. Baragetti A, Mattavelli E, Grigore L, Pellegatta F, Magni P, Catapano AL. Targeted plasma proteomics to predict the development of carotid plaques. *Stroke.* (2022) 53(9):e411–4. doi: 10.1161/strokeaha.122.038887

10. Benjamin BS, Joshua C, Matthew T, Christian B, Yi-Hsiang H, Tom GR, et al. Genetic regulation of the human plasma proteome in 54,306 UK biobank participants. *bioRxiv.* (2022). 2022.06.17.496443. doi: 10.1101/2022.06.17.496443

11. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* (2015) 12(3):e1001779. doi: 10.1371/journal.pmed.1001779

12. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. *Am J Epidemiol.* (2017) 186(9):1026–34. doi: 10.1093/aje/kwx246

13. Wik L, Nordberg N, Broberg J, Björkesten J, Assarsson E, Henriksson S, et al. Proximity extension assay in combination with next-generation sequencing for high-throughput proteome-wide analysis. *Mol Cell Proteomics.* (2021) 20:100168. doi: 10.1016/j.mcpro.2021.100168

14. Littlejohns TJ, Holliday J, Gibson LM, Garratt S, Oesingmann N, Alfaro-Almagro F, et al. The UK biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nat Commun.* (2020) 11(1):2624. doi: 10.1038/s41467-020-15948-9

15. Coffey S, Lewandowski AJ, Garratt S, Meijer R, Lynum S, Bedi R, et al. Protocol and quality assurance for carotid imaging in 100,000 participants of UK biobank: development and assessment. *Eur J Prev Cardiol.* (2017) 24(17):1799–806. doi: 10.1177/2047487317732273

16. Sinnott-Armstrong N, Tanigawa Y, Amar D, Mars N, Benner C, Aguirre M, et al. Genetics of 35 blood and urine biomarkers in the UK biobank. *Nat Genet.* (2021) 53(2):185–94. doi: 10.1038/s41588-020-00757-z

17. Thompson DJ, Wells D, Selzam S, Peneva I, Moore R, Sharp K, et al. UK biobank release and systematic evaluation of optimised polygenic risk scores for 53 diseases and quantitative traits. *medRxiv.* (2022):2022.06.16.22276246. doi: 10.1101/2022.06.16.22276246

18. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* (2005) 102(43):15545–50. doi: doi:10.1073/pnas.0506580102

19. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* (2005) 102(43):15545–50. doi: 10.1073/pnas.0506580102

20. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet.* (2003) 34(3):267–73. doi: 10.1038/ng1180

21. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics.* (2011) 27(12):1739–40. doi: 10.1093/bioinformatics/btr260

22. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* (2015) 1(6):417–25. doi: 10.1016/j.cels.2015.12.004

23. Tschiderer L, Seekircher L, Izzo R, Mancusi C, Manzi MV, Baldassarre D, et al. Association of intima-media thickness measured at the common carotid artery with incident carotid plaque: individual participant data meta-analysis of 20 prospective studies. *J Am Heart Assoc.* (2023) 12(12):e027657. doi: 10.1161/jaha.122.027657

24. Zhang W, Zhu J, Wu X, Feng T, Liao W, Li X, et al. Phenotypic and genetic effect of carotid intima-media thickness on the risk of stroke. *Hum Genet.* (2024). doi: 10.1007/s00439-024-02666-1

25. AlGhibiwi HK, Sarawi WS, Alosaimi ME, Alhusaini AM, Assiri MA, Algarzae NK. The association between cardiovascular risk factors and carotid intima-media thickness in 42,726 adults in UK biobank: a cross-sectional study. *J Cardiovasc Dev Dis.* (2023) 10(9):358. doi: 10.3390/jcdd10090358

26. Leon-Mimila P, Wang J, Huertas-Vazquez A. Relevance of multi-omics studies in cardiovascular diseases. mini review. *Front Cardiovasc Med.* (2019) 6:91. doi: 10.3389/fcvm.2019.00091

27. Corbacho-Alonso N, Baldán-Martín M, López JA, Rodríguez-Sánchez E, Martínez PJ, Mourino-Alvarez L, et al. Novel molecular plasma signatures on cardiovascular disease can stratify patients throughout life. *J Proteomics.* (2020) 222:103816. doi: 10.1016/j.jprot.2020.103816

28. Wang JG, Staessen JA, Li Y, Van Bortel LM, Nawrot T, Fagard R, et al. Carotid intima-media thickness and antihypertensive treatment: a meta-analysis of randomized controlled trials. *Stroke.* (2006) 37(7):1933–40. doi: 10.1161/01.Str.0000227223.90239.13

29. Agbaje AO, Barker AR, Tuomainen TP. Cumulative muscle mass and blood pressure but not fat mass drives arterial stiffness and carotid intima-media thickness progression in the young population and is unrelated to vascular organ damage. *Hypertens Res.* (2023) 46(4):984–99. doi: 10.1038/s41440-022-01065-1

30. Oren A, Vos LE, Uiterwaal CSPM, Grobbee DE, Bots ML. Cardiovascular risk factors and increased carotid intima-media thickness in healthy young adults: the atherosclerosis risk in young adults (ARYA) study. *Arch Intern Med*. (2003) 163 (15):1787–92. doi: 10.1001/archinte.163.15.1787

31. Moreno M, Puig J, Moreno-Navarrete JM, Xifra G, Ortega F, Ricart W, et al. Lean mass, and not fat mass, is an independent determinant of carotid intima media thickness in obese subjects. *Atherosclerosis*. (2015) 243(2):493–8. doi: 10. 1016/j.atherosclerosis.2015.09.012

32. Pillay P, Carter J, Taylor H, Lewington S, Clarke R. Independent relevance of different measures of adiposity for carotid intima-media thickness in 40 000 adults in UK biobank. *J Am Heart Assoc*. (2023) 12(2):e026694. doi: 10.1161/jaha.122.026694

33. Takase M, Nakamura T, Nakaya N, Kogure M, Hatanaka R, Nakaya K, et al. Associations between the combined fat mass Index and fat-free mass Index with carotid intima-media thickness in a Japanese population: the Tohoku medical megabank community-based cohort study. *J Atheroscler Thromb*. (2023) 30(3):255–73. doi: 10.5551/ jat.63523

34. Helle KB, Metz-Boutigue MH, Cerra MC, Angelone T. Chromogranins: from discovery to current times. *Pflugers Arch*. (2018) 470(1):143–54. doi: 10.1007/ s00424-017-2027-6

35. Madonna R, Barachini S, Ghelardoni S, Lu L, Shen WF, De Caterina R. Vasostatins: new molecular targets for atherosclerosis, post-ischaemic angiogenesis, and arteriogenesis. *Cardiovasc Res*. (2024) 120(2):132–9. doi: 10.1093/cvr/cvae008

36. Lalmanach G, Kasabova-Arjomand M, Lecaille F, Saidi A. Cystatin M/E (cystatin 6): a Janus-faced cysteine protease inhibitor with both tumor-suppressing and tumor-promoting functions. *Cancers (Basel)*. (2021) 13(8):1877. doi: 10.3390/ cancers13081877

37. Zhang Z, Zhan F. Type 2 cystatins and their roles in the regulation of human immune response and cancer progression. *Cancers (Basel)*. (2023) 15(22):5363. doi: 10.3390/cancers15225363

38. Lunde NN, Holm S, Dahl TB, Elyouncha I, Sporsheim B, Gregersen I, et al. Increased levels of legumain in plasma and plaques from patients with carotid atherosclerosis. *Atherosclerosis*. (2017) 257:216–23. doi: 10.1016/j.atherosclerosis.2016.11.026

39. Parvanova A, Reseghetti E, Abbate M, Ruggenenti P. Mechanisms and treatment of obesity-related hypertension-part 1: mechanisms. *Clin Kidney J*. (2024) 17(1): sfad282. doi: 10.1093/ckj/sfad282

40. Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, Szpyt J, et al. The BioPlex network: a systematic exploration of the human interactome. *Cell*. (2015) 162 (2):425–40. doi: 10.1016/j.cell.2015.06.043

41. Bengtsson E, Neame PJ, Heinegård D, Sommarin Y. The primary structure of a basic leucine-rich repeat protein. PRELP, found in connective tissues. *J Biol Chem*. (1995) 270(43):25639–44. doi: 10.1074/jbc.270.43.25639

42. Yamashita Y, Nakada S, Yoshihara T, Nara T, Furuya N, Miida T, et al. Perlecan, a heparan sulfate proteoglycan, regulates systemic metabolism with dynamic changes in adipose tissue and skeletal muscle. *Sci Rep*. (2018) 8(1):7766. doi: 10.1038/s41598-018-25635-x

43. GTEx Consortium. The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science*. (2020) 369(6509):1318–30. doi: 10.1126/science. aaz1776

44. The Genotype-Tissue Expression (GTEx) Project. *Gene Page—pRELP*. Cambridge, MA: Broad Institute of MIT and Harvard (2024). Available online at: https://www.gtexportal.org/home/gene/PRELP (Accessed July 08, 2024).

45. Elhadad MA, Del CG-AM, Chen CW, Neumeyer S, Delerue T, Rathmann W, et al. Plasma proteome association with coronary heart disease and carotid intima media thickness: results from the KORA F4 study. *Cardiovasc Diabetol*. (2024) 23 (1):181. doi: 10.1186/s12933-024-02274-3

46. Pietzner M, Wheeler E, Carrasco-Zanini J, Kerrison ND, Oerton E, Koprulu M, et al. Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. *Nat Commun. Nov*. (2021) 12(1):6822. doi: 10.1038/s41467-021-27164-0

47. Katz DH, Robbins JM, Deng S, Tahir UA, Bick AG, Pampana A, et al. Proteomic profiling platforms head to head: leveraging genetics and clinical traits to compare aptamer- and antibody-based methods. *Sci Adv*. (2022) 8(33):eabm5164. doi: 10. 1126/sciadv.abm5164

48. Raffield LM, Dang H, Pratte KA, Jacobson S, Gillenwater LA, Ampleford E, et al. Comparison of proteomic assessment methods in multiple cohort studies. *Proteomics*. (2020) 20(12):e1900278. doi: 10.1002/pmic.201900278

49. Eldjarn GH, Ferkingstad E, Lund SH, Helgason H, Magnusson OT, Gunnarsdottir K, et al. Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature*. (2023) 622(7982):348–58. doi: 10.1038/ s41586-023-06563-x