Journal of the American Heart Association

ORIGINAL RESEARCH

Enhanced Prediction of Peripheral Artery Disease Using Plasma Proteomics Among Individuals Without Diabetes

Meng-Yuan Miao, MD*; Jie-Qiong Lyu, MPH*; Fei Fang, BSc; Zhong-Yue Liu, BSc; Miao Xu, BSc; Hong-Jun Zhang, MD; Xing-Qiang Pan, MD; Li-Qiang Qin, PhD; Hai-Peng Wang, MD; Guo-Chong Chen, PhD

BACKGROUND: Although peripheral artery disease (PAD) is an important diabetes complication, a substantial proportion of cases occur among individuals without diabetes. This study aimed to assess the predictive value of plasma proteomics in the long-term risk of PAD among individuals initially free of diabetes.

METHODS: Included were 46 508 participants (6046 with prediabetes) without diabetes or major cardiovascular disease at recruitment of the UK Biobank. Using multivariable Cox regression models, a total of 2923 unique plasma proteins were assessed for the associations with incident PAD. Significant proteins were subsequently processed by a trained light gradient boosting machine classifier to determine important proteins. Using receiver operating characteristic analyses, the performance of these important proteins in predicting incident PAD were evaluated, in the whole sample and by glycemic status (normoglycemia and prediabetes).

RESULTS: During a median follow-up of 12.7 years, 461 participants developed PAD. There were 107 proteins associated with incident PAD, with 103 positive associations. The LGBM approach identified 9 proteins (eg, WFDC2 [WAP 4-disulfide core domain protein 2], MMP12 [macrophage metalloelastase], and GDF15 [growth differentiation factor 15]) as the top-ranked proteins based on their importance ordering. Whereas glycated hemoglobin showed very modest predictive accuracy, a panel incorporating these top proteins showed good performance in the prediction of PAD risk (area under the curve 0.820), and it significantly enhanced the prediction beyond traditional risk factors (raising area under the curve from 0.803 to 0.837, DeLong test $P=5.21\times10^{-3}$). These observations were consistent for participants with normoglycemia or prediabetes.

CONCLUSIONS: Plasma protein biomarkers enhance the prediction of long-term risk for PAD among individuals without diabetes, regardless of glycemic status.

Key Words: peripheral artery disease ■ proteomics ■ risk prediction ■ UK Biobank

Peripheral artery disease (PAD) is a major vascular condition characterized by atherosclerotic occlusion of the arteries in the lower extremities.[1] PAD is associated with higher cardiovascular morbidity and mortality regardless of the presence of symptoms, representing a considerable public health concern.[2] PAD is a complex disease with multiple pathophysiological pathways, and substantial residual risk remains that is not fully captured by the known traditional risk factors.[3] Therefore, accurate identification of novel

Correspondence to: Guo-Chong Chen, PhD, Department of Cardiology, The First Affiliated Hospital, School of Public Health, Suzhou Medical College of Soochow University, 199 Ren'ai Road, Suzhou 215123, China. Email: gcchen@suda.edu.cn and Hai-Peng Wang, MD, Department of Cardiology, The First Affiliated Hospital of Soochow University, 188 Shizi Street, Suzhou 215006, China. Email: wanghaipeng0324@suda.edu.cn

*M.-Y. Miao and J.-Q. Lyu are co-first authors.

This article was sent to Jacquelyn Y. Taylor, PhD, PNP-BC, RN, FAHA, FAAN, Associate Editor, for review by expert referees, editorial decision, and final disposition.

Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.125.047797

For Sources of Funding and Disclosures, see page 9.

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

CLINICAL PERSPECTIVE

What Is New?

- In this large-scale plasma proteomics study of individuals without diabetes, 107 plasma proteins were independently associated with incident peripheral artery disease, implicating important inflammatory and immune-related pathways.
- A light gradient boosting machine approach identified 9 top-ranked proteins, and these plasma protein biomarkers enhance the prediction of long-term risk for peripheral artery disease among individuals without diabetes, regardless of glycemic status.

What Are the Clinical Implications?

- The identified protein panel may facilitate the early identification of at-risk individuals and better prediction of peripheral artery disease risk, especially for those without diabetes who are not regularly screened for peripheral artery disease.

Nonstandard Abbreviations and Acronyms

GDF15	growth differentiation factor 15
MMP12	macrophage metalloelastase
UKB	UK Biobank
WFDC2	WAP 4-disulfide core domain protein 2

molecular markers is highly desirable for better refining risk prediction and understanding the pathogenesis of PAD.

The rapid advancement of proteomics technologies has significantly accelerated the discovery of novel disease biomarkers.[4] Recent proteomics studies have proposed that models incorporating protein biomarkers produced superior predictions of multiple health outcomes, such as cardiovascular diseases[5–7] and diabetes.[8,9] PAD is among the major macrovascular complications of diabetes, and a recent study demonstrated the utility of plasma proteomics in predicting risk of PAD among individuals with type 2 diabetes.[10] Nevertheless, a substantial proportion of PAD cases occur among individuals without diabetes,[11] and it has been implicated that protein-based disease risk prediction may be affected by glycemic status.[12] To this point, the value of plasma proteins in the prediction of the long-term risk of PAD remains unclear for individuals without diabetes, especially for those with normoglycemia who are unlikely to be screened regularly for the risk of developing PAD.

To fill these knowledge gaps, we conducted a series of analyses among individuals without diabetes, using data from the UKB-PPP (UK Biobank Pharma Proteomics Project), a large-scale proteomic investigation that measured 2941 plasma proteins among >50 000 individuals. First, we performed a proteome-wide association analysis to identify plasma protein biomarkers for incident PAD. Second, we determined the relative importance of PAD-associated proteins in risk prediction of PAD. Third, we assessed the predictive performance of the top-ranked proteins for incident PAD, in the whole study sample and by glycemic status (normoglycemia and prediabetes).

METHODS

The data that support the findings of this study are available upon application to the UKB (www.ukbiobank.ac.uk/).

Study Population

The UKB is a prospective population-based cohort that recruited >500 000 participants aged 37 to 73 years from 22 assessment centers across England, Scotland, and Wales. From 2006 to 2010, participants completed various questionnaires, underwent a range of physical measures, and provided biological samples. The study was approved by the Northwest Multi-Centre Research Ethics Committee (REC reference for UK Biobank 11/NW/0382) and all study participants provided written informed consent.

The UKB-PPP is a precompetitive consortium of 13 biopharmaceutical companies funding the generation of blood-based proteomic data in a subset of UKB participants.[13] In our study, among 502 411 participants from UKB, a subset of randomly selected 53 026 participants had information on 2941 protein analytes (reflecting 2923 unique proteins) measured across Olink panels in plasma samples collected at baseline. We excluded participants with prevalent diabetes, PAD, or other major cardiovascular diseases at baseline (n=6518), leaving 46 508 participants in the present analysis. Among these, there were 6046 individuals with prediabetes (39 mmol/mol ≤ glycated hemoglobin [HbA1c] <48 mmol/mol) and 40 462 with normoglycemia (HbA1c <39 mmol/mol).

Measurements of Plasma Proteomics

The baseline blood samples donated by UKB-PPP study participants underwent proteomic profiling using the Olink Explore 3072 platform (Olink Proteomics), which quantified 2941 protein analytes representing 2923 unique proteins across 8 panels (Cardiometabolic, Cardiometabolic II, Inflammation, Inflammation II, Neurology, Neurology II, Oncology, and Oncology II). The details on the sample collection, plasma sample preparation, plasma profiling

using the Olink technology, and quality control were described elsewhere.[13,14] There were 830 proteins with ≥25% of their data below the lowest level of detection or with >50% of data failing quality control, which were excluded from the analysis. For the remaining 2093 proteins included, missing values were imputed by half of the lowest values detected. Data were then standardized using rank-based inverse normal transformation followed by *Z* score normalization to ensure comparability and standardization across the proteins.

Ascertainment of Incident PAD

Incident cases of PAD during the follow-up were ascertained from linked hospital and death databases. Date and cause of hospital admissions were identified via record linkage to Health Episode Statistics (England and Wales) and the Scottish Morbidity Records (Scotland). Date and cause of death were obtained from death certificates held by the National Health Service Information Centre (England and Wales) and the National Health Service Central Register Scotland (Scotland). PAD was defined using *the International Classification of Diseases, Ninth Revision* (*ICD-9*) codes, *Tenth Revision* (*ICD-10*) codes, and the Office of Population Censuses and Surveys Classification of Interventions and Procedures, version 4 codes. A complete list of these codes is provided in Table S1.

Covariates

Information on sociodemographic characteristics, lifestyle factors, medical history, and medication use was assessed using touchscreen-based questionnaires. The Townsend deprivation index was derived to capture area deprivation, with a higher score indicating a higher level of social deprivation. Physical activity was assessed using a self-reported short-form international physical activity questionnaire. A healthy diet score was calculated based on the following 6 food groups: vegetables and fruit, red meat, processed meat, whole grains, refined grains, and fish.[15] Blood pressure was measured using the Omron HEM-7015IT digital blood pressure monitor and calculated as the mean of 2 sitting measures. Total cholesterol was measured using the Beckman Coulter AU580. HbA1c assay was performed using a high-performance liquid chromatography method (Bio-Rad Variant II Turbo analyzer, Bio-Rad Laboratories). Diabetes was defined as self-reported physician's diagnosis, medication use, or HbA1c ≥48 mmol/mol.

Statistical Analysis

Baseline Characteristics of the Study Population

Baseline participant characteristics were presented according to the status of incident PAD, as mean±SD or frequencies (%) where appropriate.

Proteomic Association Analyses

Multivariable Cox proportional hazards regression models were used to estimate hazard ratios (HRs) and 95% CIs for the associations between individual plasma proteins and incident PAD. Follow-up period was calculated from the date of baseline recruitment to the date of PAD diagnosis, death, or the most recent follow-up for the approved data set (December 2021), whichever occurred first. The multivariable model was adjusted for age, sex, race/ethnicity, Townsend deprivation index, body mass index, waist circumference, tobacco consumption, alcohol consumption, total physical activity, healthy diet score, systolic blood pressure, total cholesterol, HbA1c, and use of antihypertensive or lipid-lowering medications. Missing data for continuous covariates were addressed using sex-specific median values, and categorical covariates were addressed with a missing indicator category.[16] Missingness in continuous covariates was generally low, with most variables showing missingness proportions below 5%, except for HbA1c (5.42%), systolic blood pressure (6.18%), and physical activity (19.41%). To characterize the missing-data pattern, missingness indicators were modeled against observed baseline participants characteristics as well as outcome information using logistic regression. The results suggested that the missing-data pattern was unlikely to be strictly missing completely at random and may be more consistent with a missing-at-random mechanism (Table S2). To evaluate the potential impact of simple missing-data handling on the main analyses, an additional sensitivity analysis was conducted by restricting the sample to participants with complete data on all continuous covariates and repeating the primary analyses. Bonferroni corrections were applied to assess significant associations ($P<0.05$), accounting for the number of proteins tested ($k=2093$).

Pathway Enrichment Analyses

To gain insights into the potential biological pathways for the significant proteins, we performed systematic enrichment analyses encompassing Gene Ontology, Kyoto Encyclopedia of Genes and Genomes, and Reactome pathway analyses using the R package *clusterProfiler*. Statistical significance was presented as the *P* value of Fisher's exact test, followed by false discovery rate corrections via the Benjamini–Hochberg procedure.

Proteins Importance Ranking

The relative importance of the PAD-related proteins in the risk prediction of PAD were determined mainly by 2 steps: candidate proteins ranking and sequential forward selection. In the first step, the significant proteins

identified from the association analyses were fed into a preliminary trained light gradient boosting machine classifier, and each protein was ranked according to its contribution to model performance based on the information gain, which can be regarded as the predictor's ability to identify future incident PAD. Afterwards, a sequential forward selection strategy was applied, where predictors were added one by one to the newly developed light gradient boosting machine classifier according to the predictor importance ranking. This iterative process continued until the optimal performance in terms of the area under the curve (AUC) was reached, determined when the inclusion of an additional predictor failed to yield a statistically significant improvement in model performance, as assessed by 2 consecutive DeLong tests.

Evaluation of Model Performance in Predicting PAD

Subsequently, we used the top-ranked proteins to construct following prediction models with or without plasma proteins, including (1) a single important protein, (2) traditional risk factors, (3) HbA1c, (4) a protein panel including all important proteins, and (5) the protein panel + traditional risk factors. Traditional risk factors consisted of all the covariates included in the aforementioned multivariable models. HbA1c is a well-established clinical indicator for assessing glucose metabolism, widely used to guide diabetes management and evaluate vascular risk indirectly. Treating HbA1c as a standalone model could allow us to directly compare the predictive performance of protein panel against a marker routinely measured in clinical practice. Receiver operating characteristic analyses were performed to evaluate the performance of the models in predicting PAD. DeLong tests and bootstrap tests with 2000 iterations were adopted to compare whether AUC values differed significantly between models. The data were split into 10 folds based on the geographical locations (East Midlands, London, North East, North West, Scotland, South East, South West, Wales, West Midlands, and Yorkshire and Humber) of participants' recruitment centers. Models were developed and validated using a 10-fold cross-validation strategy that the validation set (1-fold of data) was kept untouched and merely used for evaluation purposes, and the hyperparameters tuning and postcalibration were performed within an inner-looped cross-validation within the training sets (9-fold of data). This process was repeated 10 times, with the folds being shifted for each iteration to serve as training and testing sets. In addition, differences in the expression levels of top-ranked proteins between individuals with normoglycemia and prediabetes were assessed using 2-sample *t* tests. To evaluate the generalizability of the constructed protein-based prediction models and the robustness of their predictive accuracy, we repeated the receiver operating characteristic analysis across subgroups defined by glycemic status (normoglycemia and prediabetes).

All statistical tests were performed using Python (version 3.9) and R software (version 4.4.2, R Development Core Team). Statistical significance was set at 0.05 (2 tailed) unless otherwise stated.

RESULTS

Baseline Characteristics of the Study Population

Mean age for the 46 508 participants was 56.28 years at baseline and 20 341 (43.7%) were male. As shown in Table S3, compared with participants without PAD, those with PAD tended to be older, were more likely to be male, and had a higher Townsend deprivation index. Body mass index, waist circumference, systolic blood pressure, and glycated hemoglobin levels were higher in the group with PAD, whereas they had slightly lower total cholesterol levels. Those with PAD also were more likely to have an unhealthy lifestyle and to use medications for metabolic dysfunction.

Associations of Plasma Proteins With Incident PAD

During a median of 12.7 years of follow up, 461 participants experienced incident PAD, among whom 132 participants had prediabetes at baseline. Of the 2093 unique proteins assessed, after the multivariate adjustment, 107 (103 positively and 4 inversely) were significantly associated with incident PAD (Figure 1; Table S4). By HR, GDF15 (growth differentiation factor 15) showed the strongest positive association and OMG (oligodendrocyte-myelin glycoprotein) had the strongest inverse association with incident PAD, with HRs of 1.67 (95% CI, 1.46–1.90; $P_{\text{Bonferroni-corrected}}=5.21\times10^{-11}$) and 0.77 (95% CI, 0.70–0.85; $P_{\text{Bonferroni-corrected}}=1.78\times10^{-4}$), respectively.

Enrichment analyses implicated multiple biological pathways related to these significant proteins, such as immune response and cell signaling (eg, cytokine-cytokine receptor interaction, chemokine signaling pathway, signaling by interleukins), cell migration and differentiation (eg, leukocyte migration, osteoclast differentiation), cell structure and extracellular matrix (eg, collagen-containing extracellular matrix, extracellular matrix organization), metabolism and transport (eg, carbohydrate binding), and signal transduction and gene expression (eg, positive regulation of MAPK [mitogen-activated protein kinase] cascade) (Figure S1).

Protein Importance Ranking

On the basis of the 107 PAD-related proteins, we further sorted their relative importance in the prediction task

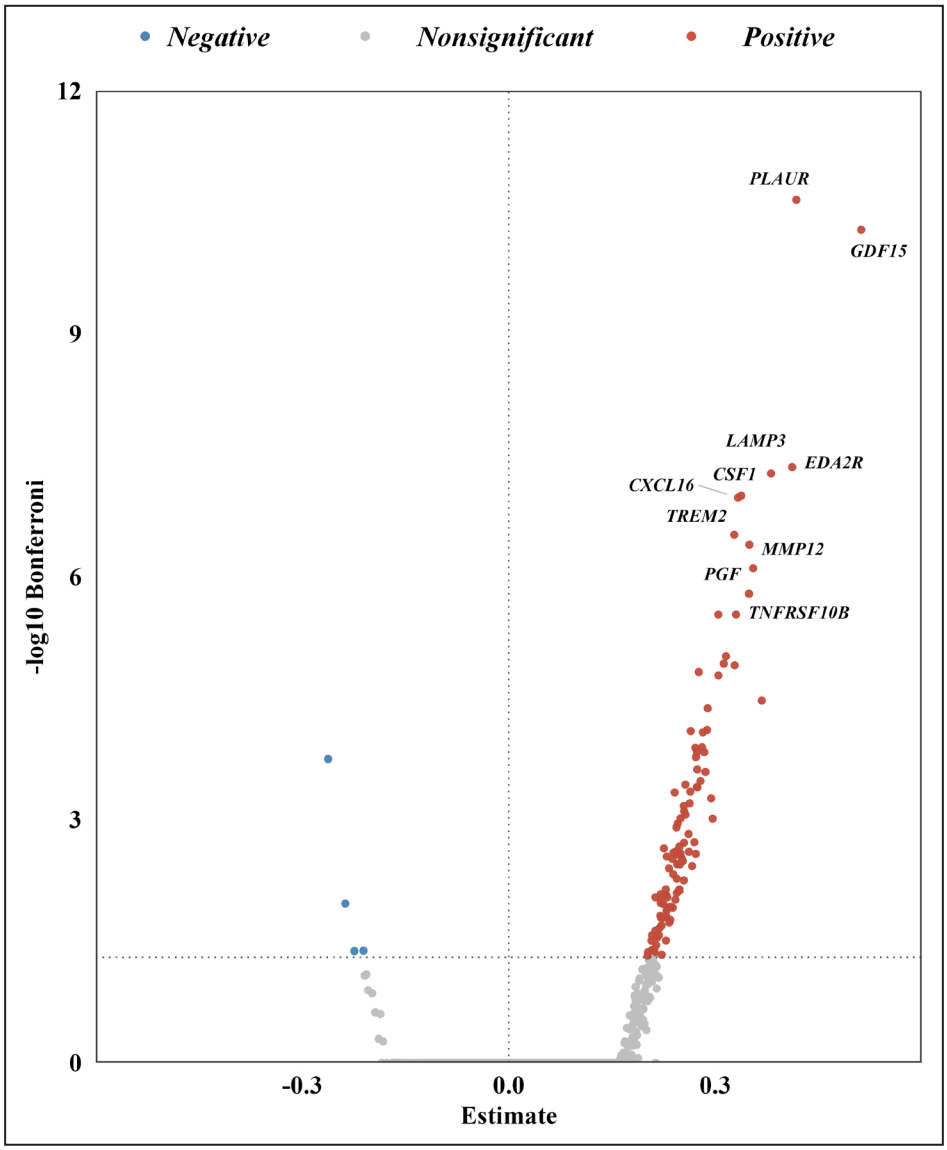

Figure 1. Associations of plasma proteins with incident PAD among individuals without diabetes.

Volcano plots showing the multivariable-adjusted beta coefficients (*x* axis) and $-\log_{10}$ Bonferroni (*y* axis) for the associations of 2093 proteins with incident PAD. Results were from Cox proportional hazard regression models adjusted for age, sex, race, ethnicity, Townsend deprivation index, body mass index, waist circumference, tobacco consumption, alcohol consumption, total physical activity, healthy diet score, systolic blood pressure, total cholesterol, glycated hemoglobin, and medication history (antihypertensive, and lipid lowering). Proteins above the horizontal dotted black line were significantly associated with incident PAD after Bonferroni corrections ($P<0.05$) taking into account the number of proteins tested (k=2093). CSF1 indicates colony-stimulating factor 1; CXCL16, C-X-C motif chemokine 16; EDA2R, tumor necrosis factor receptor superfamily member 27; GDF15, growth differentiation factor 15; LAMP3, lysosome-associated membrane glycoprotein 3; MMP12, macrophage metalloelastase; PGF, placental growth factor; PLAUR, urokinase plasminogen activator surface receptor; TNFRSF10B, tumor necrosis factor receptor superfamily member 10b; and TREM2, triggering receptor expressed on myeloid cells 2.

(Table S5). The predictive importance ranking for these proteins is illustrated in Figure S2, and the top 30 proteins are detailed in Figure 2A. The predictive capacity for PAD (quantified by AUC) demonstrated rapid enhancement with the initial incorporation of a few top-ranked proteins, eventually reaching a performance plateau as additional biomarkers were included. Based on sequential forward selection with statistical significance validation, 9 proteins, including WFDC2 (WAP 4-disulfide core domain protein 2), MMP12 (macrophage metalloelastase),

GDF15, CXCL17 (C-X-C motif chemokine 17), LAMP3 (lysosome-associated membrane glycoprotein 3), LRRN1 (leucine-rich repeat neuronal protein 1), BCAN (brevican core protein), EDA2R (tumor necrosis factor receptor superfamily member 27), and NEFL (neurofilament light polypeptide), were ultimately determined as important proteins for risk prediction of PAD. The associations between these 9 key proteins and incident PAD are presented in Figure 2B.

Predictive Performance of Plasma Proteins for Incident PAD

We then examined the predictive accuracy of the selected important proteins for long-term risk of PAD, using a 10-fold cross-validation approach. As shown in Table S6, the top 3 important proteins, WFDC2, MMP12, and GDF15, achieved predictive accuracies of 0.743, 0.751, and 0.767, respectively, which largely surpassed the AUC of HbA1c (0.631). Combining these 9 proteins as a protein panel yielded an AUC of 0.820. Adding the protein panel to traditional risk factors attained the best-performing model for incident PAD, increasing the AUC from 0.803 to 0.837 (DeLong test $P=5.21\times10^{-3}$) (Figure 2C; Table S7).

Seven of the 9 top-ranked proteins exhibited higher expression levels in the group with prediabetes than in the group with normoglycemia, whereas BCAN and LRRN1 levels were higher in the group with normoglycemia (Figure S3). Despite these glycemic-related expression heterogeneities, the protein panel maintained robust predictive performance and demonstrated strong predictive ability for incident PAD across the subgroups defined by glycemic status (Figure 3; Table S8), with AUC of 0.807 in normoglycemia and 0.792 in prediabetes, both considerably outperforming the AUC of HbA1c (0.555 in normoglycemia and 0.550 in prediabetes). For both subgroups, adding the protein panel substantially improved the prediction accuracy beyond HbA1c and other traditional risk factors (all *P* values from DeLong test <0.05) (Table S9), with AUC increasing from 0.790 to 0.830 in normoglycemia and from 0.721 to 0.795 in prediabetes (Figure 3).

Among 2733 participants who had diabetes and were free of PAD at baseline, we conducted a supplementary analysis to assess the performance of the 9-protein panel in predicting PAD risk (115 incident cases). In this group, the protein panel achieved an AUC of 0.718, substantially lower than that for individuals without diabetes (AUC=0.820), suggesting that the identified proteins may capture biological processes beyond the progression of diabetes.

Sensitivity Analysis

After excluding participants with missingness in continuous covariates, the overall association estimates were highly concordant with those in the primary analysis (Figure S4). Although the number of significant proteins decreased to 26, all these proteins were among the 107 significant proteins identified in the primary analysis. Of the 9 top-ranked proteins, all but 2 proteins (LRRN1 and BCAN) remained significantly associated with incident PAD after Bonferroni correlation (Table S10). A 7-protein panel still showed good predictive performance for incident PAD and significantly improved discrimination beyond the basic model, regardless of glycemic status (Figure S5).

DISCUSSION

Although PAD is an important macrovascular complication of diabetes, it also occurs among individuals without diabetes. Yet, there has been limited epidemiological evidence regarding the risk prediction of PAD for individuals without diabetes who are less likely to be screened for PAD. In this large-scale plasma proteomics study of individuals without diabetes, we identified 107 significant protein-PAD associations potentially involving important inflammatory and immune-related processes. Among these, 9 proteins (eg, WFDC2, MMP12, and GDF15) were identified as the most important proteins for predicting the long-term risk of PAD. Although HbA1c showed very modest predictive accuracy, a panel of these key proteins showed good performance in predicting the risk of PAD, and it substantially improved the risk prediction beyond HbA1c and other traditional risk factors, regardless of glycemic status.

Recent proteome studies have highlighted the potential of proteomic biomarkers to improve the prediction of vascular conditions.[17] For instance, in the Second Manifestations of Arterial Disease cohort, incorporating 50 proteins into a cardiovascular risk model enhanced the ability of predicting recurrent atherosclerotic cardiovascular disease.[18] In the Heart and Soul study, a risk score based on 9 proteins performed better than the refit Framingham secondary event risk score in predicting cardiovascular events among patients with stable coronary heart disease.[19] PAD shares many risk factors with other vascular diseases,[20] indicating potential value of proteomic biomarkers in the prediction of PAD risk. Although previous studies have linked proteomic biomarkers to the development or prediction of PAD among individuals with diabetes,[10] relevant evidence remains limited for those without diabetes. In the present study of participants without diabetes, we identified a set of proteins associated with incident PAD. Several of these individual proteins, such as GDF15, PLAUR (urokinase plasminogen activator surface receptor), and MMP12, which are among the top 10 significant proteins in the present study, have previously been

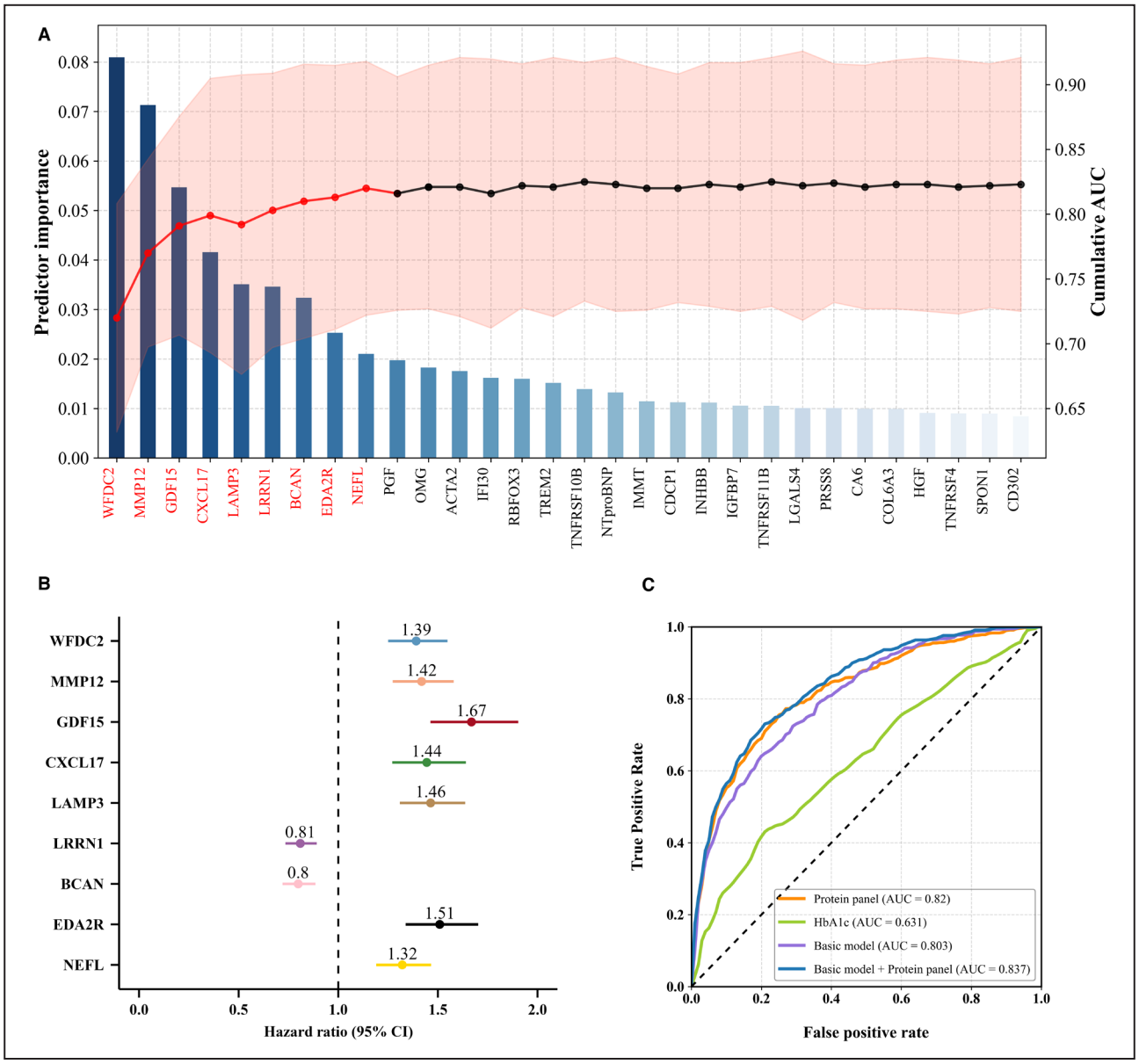

Figure 2. Protein importance ranking and predictive performance of plasma proteins for incident PAD among individuals without diabetes.

A, Sequential forward selection from the plasma proteins significantly associated with incident PAD (k=107). Ranking of predictive importance for the top 30 proteins is shown in the figure and that for all 107 proteins is presented in Figure S2. The bar chart (**left** axis) indicates the importance of the sorted proteins based on their contributions to the prediction of future PAD (as judged by the information gain). The line chart (**right** axis) illustrates cumulative AUC values upon the inclusion of proteins one by one in each iteration. The top-ranked key proteins are marked in red. Shaded regions represent SEs derived from cross-validation. **B**, The forest plot shows the hazard ratios (95% CIs) for the associations between the selected top proteins and risk of PAD. **C**, Receiver operating curves show the performance of different variable models for predicting incident PAD in individuals without diabetes. Protein panel refers to a combination of 9 top proteins selected by light gradient boosting machine classifier. Basic model included all covariates listed in the legend for Figure 1. ACTA2 indicates actin alpha 2, smooth muscle; AUC, area under the curve; BCAN, brevican core protein; CA6, carbonic anhydrase 6; CDCP1, CUB domain-containing protein 1; COL6A3, collagen type VI alpha 3 chain; CXCL17, C-X-C motif chemokine 17; EDA2R, tumor necrosis factor receptor superfamily member 27; GDF15, growth differentiation factor 15; HbA1c, glycated hemoglobin; HGF, hepatocyte growth factor; IFI30, gamma-interferon–inducible lysosomal thiol reductase; IGFBP7, insulin-like growth factor binding protein 7; IMMT, mitochondrial inner membrane protein; INHBB, inhibin subunit beta B; LAMP3, lysosome-associated membrane glycoprotein 3; LGALS4, galectin-4; LRRN1, leucine-rich repeat neuronal protein 1; MMP12, macrophage metalloelastase; NEFL, neurofilament light polypeptide; NT-proBNP, N-terminal pro-B-type natriuretic peptide; OMG, oligodendrocyte-myelin glycoprotein; PGF, placental growth factor; PLAUR, urokinase plasminogen activator surface receptor; PRSS8, prostatin; RBFOX3, RNA binding protein fox-1 homolog 3; SPON1, spondin-1; TNFRSF10B, tumor necrosis factor receptor superfamily member 10b; TNFRSF11B, tumor necrosis factor receptor superfamily member 11b; TREM2, triggering receptor expressed on myeloid cells 2; and WFDC2, WAP four-disulfide core domain protein 2.

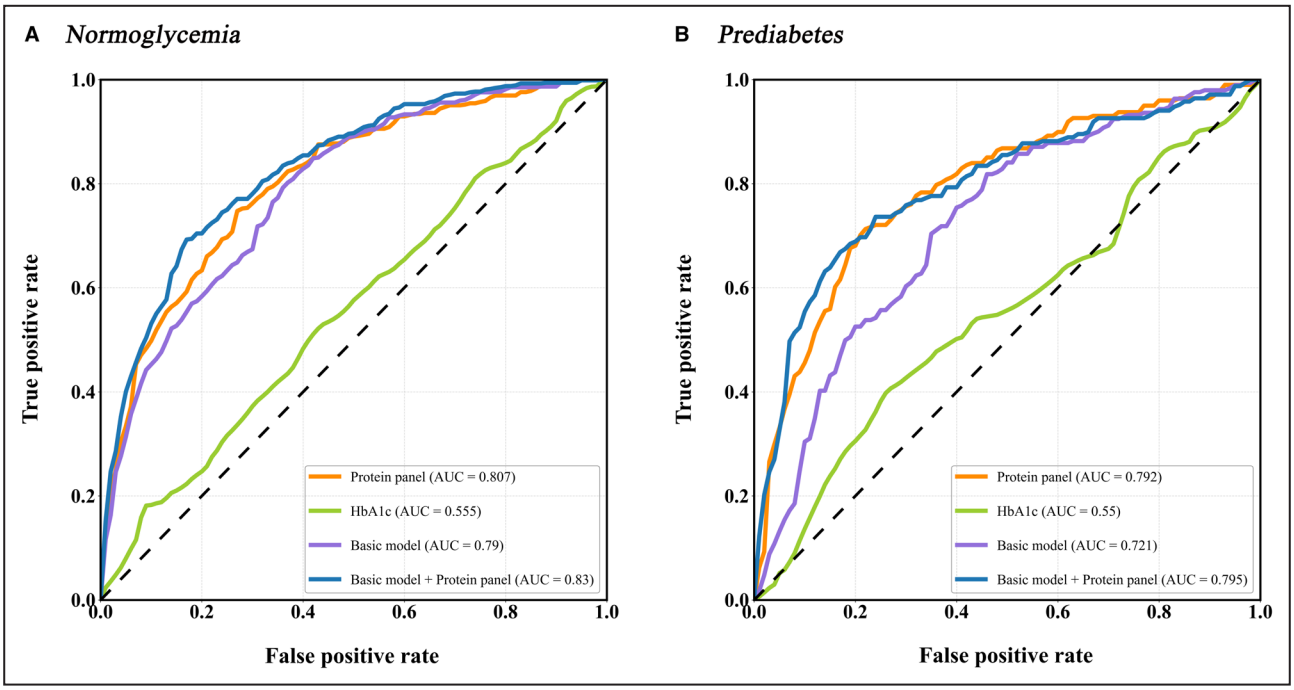

Figure 3. Predictive performance of plasma proteins for incident PAD in normoglycemia and prediabetes.
Receiver operating curves showed the performance of different variable models for predicting incident PAD in (**A**) group with normoglycemia group, and (**B**) group with prediabetes group. Protein panel refers to a combination of the 9 top proteins selected by light gradient boosting machine classifier. Basic model included all covariates listed in the legend for Figure 1. AUC indicates area under the curve; and HbA1c, glycated hemoglobin.

proposed as biomarkers for PAD in the general population.[21,22] In addition, we found that the identified PAD-related proteins shared important pathways that have been implicated in the pathogenesis of atherosclerotic cardiovascular diseases,[23–25] such as immune response and cell signaling, cell structure and extracellular matrix, metabolism and transport, and signal transduction and gene expression.

Our findings showed the greatest importance of 9 plasma proteins in predicting incident PAD, with WFDC2, MMP12, and GDF15 ranking as the top 3 proteins substantially contributing to the predictive ability (combined AUC for the 3 proteins: 0.791). WFDC2, also known as human epididymal protein 4, functions as an inhibitor of multiple proteinases, playing a crucial role in modulating the balance between extracellular matrix synthesis and degradation, as well as regulating inflammatory responses.[26] These processes are essential in the development of atherosclerotic vascular disease and the maintenance of plaque stability.[27] WFDC2 has been reported to be associated with cardiovascular events,[26] but its relationship with PAD has not been studied. The positive relationship of plasma WFDC2 with incident PAD, along with its potential as the strongest predictor for PAD, are our novel findings that need further confirmation in future studies. MMP12, a matrix metalloproteinase zinc-dependent proteolytic enzyme secreted from immunocompetent cells, has been associated with several atherosclerotic cardiovascular conditions, such as intima-media thickness in the common carotid artery,[28,29] coronary events,[30] and large-artery atherosclerotic stroke.[31] Results from a Swedish cohort study showed that higher MMP12 levels were associated with an increased risk of PAD in the general population.[21] GDF15 is a member of the transforming growth factor family involved in apoptosis and inflammation, and it has been found to be differentially expressed in patients with and without PAD.[32] Higher levels of GDF15 have been associated with an increased risk of PAD in the general population as well as in individuals with diabetes.[21,33]

HbA1c is commonly used as a clinical marker to guide the management of diabetes and its prognosis. Nevertheless, given the multifactorial nature of PAD pathogenesis and the heterogeneity of clinical manifestations,[34] a single biomarker is unlikely to attain the highest predictive accuracy. In the present study of individuals without diabetes, HbA1c showed only modest ability in predicting the risk of PAD, highlighting the need to discover other biomarkers, alongside phenotypic and clinical features, to improve early risk screening and disease prevention.[32] In our study, a panel of the 9 top-ranked proteins demonstrated good predictive performance for incident PAD and significantly enhanced the risk discrimination beyond HbA1c and

other traditional risk factors. We further found that the predictive performance of the protein panel was similar for individuals with normoglycemia and those with prediabetes. In a previous proteomic analysis of 354 patients undergoing peripheral or coronary angiography, a protein panel showed similar performance in the prediction of obstructive PAD across HbA1c levels.[35] Our study used a larger sample size with a greater number of protein biomarkers, extending the prior knowledge to individuals with normoglycemia or prediabetes.

Findings from this large-scale, machine learning-driven proteome study of individuals without diabetes may have important significance both for public health and clinical practices. The selected top-ranked protein biomarkers may shed novel mechanistic insight into the pathological process of PAD that is independent of glucose metabolism. The identified protein panel may facilitate the early identification of at-risk individuals and better prediction of PAD risk, especially for those without diabetes who are not regularly screened for PAD.

Our study benefits from the large sample size, the long-term follow up, and the high-throughput proteome analysis, although several potential limitations should be acknowledged. Although a comprehensive assessment of plasma proteins is provided by the UKB-PPP, there are some proteins that are not captured by the platform. In addition, independent external cohorts with large-scale proteomics data are unavailable to validate our findings. In the present study, we have implemented a rigorous 10-fold cross-validation strategy and confirmed the model performance in predefined subgroups by glycemic status (normoglycemia and prediabetes). This rigorous internal validation is widely used for ensuring model robustness and generalizability,[17,36,37] which align with recent proteomic studies for biomarker discovery. Moreover, participants in the UKB were predominantly White Europeans, and our findings are yet to be confirmed for other regional or ethnic populations.

CONCLUSIONS

In summary, our large-scale proteomic analysis identified a group of plasma proteins associated with incident PAD among individuals without diabetes, emphasizing the importance of 9 key proteins (especially WFDC2, MMP12, and GDF15) in predicting the long-term risk for PAD beyond traditional risk factors including glycemic levels. These findings may provide new insights into the disease mechanisms that are beyond glucose metabolism. Our findings may also have implications for early screening and targeted prevention of PAD among the general population who are unlikely to be regularly screened for the risk of PAD.

ARTICLE INFORMATION

Received November 7, 2025; accepted April 2, 2026.

Affiliations

Department of Cardiology, The First Affiliated Hospital, School of Public Health, Suzhou Medical College of Soochow University, Suzhou, China (M-Y.M., J-Q.L., F.F., Z-Y.L., M.X., L-Q.Q., H-P.W., G-C.C.); MOE Key Laboratory of Geriatric Diseases and Immunology, Suzhou Medical College, Soochow University, Suzhou, China (M-Y.M., G-C.C.); Yancheng Center for Disease Control and Prevention, Yancheng, China (M.X., H-J.Z.); and Ningbo Municipal Center for Disease Control and Prevention, Ningbo, China (X-Q.P.).

Acknowledgments

The authors thank the UK Biobank participants. This research has been conducted using the UK Biobank Resource under Application Number 90087. This work uses data provided by patients and collected by the National Health Service as part of their care and support. All software used in this study is publicly available. The code used in this study can be accessed at https://github.com/MMYBBYXY/PAD_Proteomics_Prediction. Meng-Yuan Miao: Visualization, formal analysis, writing—original draft, writing—review and editing. Jie-Qiong Lyu: Visualization, formal analysis, writing—review and editing. Fei Fang: Formal analysis. Zhong-Yue Liu: Formal analysis. Miao Xu: Writing—review and editing. Hong-Jun Zhang: Writing—review and editing. Xing-Qiang Pan: Writing—review and editing. Li-Qiang Qin: Writing—review and editing. Hai-Peng Wang: Conceptualization, visualization, writing—review and editing. Guo-Chong Chen: Conceptualization, supervision, visualization, writing—review and editing.

Sources of Funding

This study is partially supported by the Gusu Leading Talent Plan for Scientific and Technological Innovation and Entrepreneurship (ZXL2023345) and the Priority Academic Program Development of Jiangsu Higher Education Institutions. Dr Pan is supported by Major Project of Scientific and Technological Innovation 2035 of Ningbo Science and Technology Bureau (2024Z234) and Ningbo Science and Technology Development Special Project Key Research and Development Plan for 2023 (2023Z183).

Disclosures

None.

Supplemental Material

Tables S1–S10
Figures S1–S5

REFERENCES

1. Hirsch AT, Haskal ZJ, Hertzer NR, Bakal CW, Creager MA, Halperin JL, Hiratzka LF, Murphy WR, Olin JW, Puschett JB, et al. ACC/AHA 2005 practice guidelines for the management of patients with peripheral arterial disease (lower extremity, renal, mesenteric, and abdominal aortic): a collaborative report from the American Association for Vascular Surgery/Society for Vascular Surgery, Society for Cardiovascular Angiography and Interventions, Society for Vascular Medicine and Biology, Society of Interventional Radiology, and the ACC/AHA task force on practice guidelines (writing committee to develop guidelines for the management of patients with peripheral arterial disease): endorsed by the American Association of Cardiovascular and Pulmonary Rehabilitation; National Heart, Lung, and Blood Institute; Society for Vascular Nursing; TransAtlantic Inter-Society Consensus; and Vascular Disease Foundation. *Circulation*. 2006;113:e463–e654. doi: 10.1161/CIRCULATIONAHA.106.173994
2. Criqui MH, Langer RD, Fronek A, Feigelson HS, Klauber MR, McCann TJ, Browner D. Mortality over a period of 10 years in patients with peripheral arterial disease. *N Engl J Med*. 1992;326:381–386. doi: 10.1056/NEJM199202063260605
3. Zierfuss B, Höbaus C, Herz CT, Koppensteiner R, Stangl H, Schernthaner GH. High-density lipoprotein particle subclasses in statin-treated patients with peripheral artery disease predict long-term survival. *Thromb Haemost*. 2022;122:1804–1813. doi: 10.1055/a-1827-7896

4. Smith JG, Gerszten RE. Emerging affinity-based proteomic technologies for large-scale plasma profiling in cardiovascular disease. *Circulation*. 2017;135:1651–1664. doi: 10.1161/CIRCULATIONAHA.116.025446
5. Ho FK, Mark PB, Lees JS, Pell JP, Strawbridge RJ, Kimenai DM, Mills NL, Woodward M, McMurray JJV, Sattar N, et al. A proteomics-based approach for prediction of different cardiovascular diseases and dementia. *Circulation*. 2024;151:277–287. doi: 10.1161/CIRCULATIONAHA.124.070454
6. Gan X, Yang S, Zhang Y, Ye Z, Zhang Y, Xiang H, Huang Y, Wu Y, Zhang Y, Qin X. Large-scale plasma proteomics profiles for predicting ischemic stroke risk in the general population. *Stroke*. 2024;56:456–464. doi: 10.1161/STROKEAHA.124.048654
7. Schuermans A, Pournamdari AB, Lee J, Bhukar R, Ganesh S, Darosa N, Small AM, Yu Z, Hornsby W, Koyama S, et al. Integrative proteomic analyses across common cardiac diseases yield mechanistic insights and enhanced prediction. *Nat Cardiovasc Res*. 2024;3:1516–1530. doi: 10.1038/s44161-024-00567-0
8. Gummesson A, Björnson E, Fagerberg L, Zhong W, Tebani A, Edfors F, Schmidt C, Lundqvist A, Adiels M, Bäckhed F, et al. Longitudinal plasma protein profiling of newly diagnosed type 2 diabetes. *EBioMedicine*. 2021;63:103147. doi: 10.1016/j.ebiom.2020.103147
9. Chen ZZ, Gerszten RE. Metabolomics and proteomics in type 2 diabetes. *Circ Res*. 2020;126:1613–1627. doi: 10.1161/CIRCRESAHA.120.315898
10. Yu H, Zhang J, Qian F, Yao P, Xu K, Wu P, Li R, Qiu Z, Li R, Zhu K, et al. Large-scale plasma proteomics improves prediction of peripheral artery disease in individuals with type 2 diabetes: a prospective cohort study. *Diabetes Care*. 2025;48:381–389. doi: 10.2337/dc24-1696
11. Criqui MH, Aboyans V. Epidemiology of peripheral artery disease. *Circ Res*. 2015;116:1509–1526. doi: 10.1161/CIRCRESAHA.116.303849
12. Verdonschot JAJ, Ferreira JP, Pellicori P, Brunner-La Rocca HP, Clark AL, Cosmi F, Cuthbert J, Girerd N, Mariottoni B, Petutschnigg J, et al. Proteomic mechanistic profile of patients with diabetes at risk of developing heart failure: insights from the HOMAGE trial. *Cardiovasc Diabetol*. 2021;20:163. doi: 10.1186/s12933-021-01357-9
13. Sun BB, Chiou J, Traylor M, Benner C, Hsu YH, Richardson TG, Surendran P, Mahajan A, Robins C, Vasquez-Grinnell SG, et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature*. 2023;622:329–338. doi: 10.1038/s41586-023-06592-6
14. Eldjarn GH, Ferkingstad E, Lund SH, Helgason H, Magnusson OT, Gunnarsdottir K, Olafsdottir TA, Halldorsson BV, Olason PI, Zink F, et al. Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature*. 2023;622:348–358. doi: 10.1038/s41586-023-06563-x
15. Zhu Z, Li FR, Jia Y, Li Y, Guo D, Chen J, Tian H, Yang J, Yang HH, Chen LH, et al. Association of lifestyle with Incidence of heart failure according to metabolic and genetic risk status: a population-based prospective study. *Circ Heart Fail*. 2022;15:e009592. doi: 10.1161/CIRCHEARTFAILURE.122.009592
16. Lyu JQ, Miao MY, Wang JM, Qian YW, Han WW, Peng XZ, Tao HW, Yang J, Chen JS, Qin LQ, et al. Consumption of total and specific alcoholic beverages and long-term risk of gout among men and women. *JAMA Netw Open*. 2024;7:e2430700. doi: 10.1001/jamanetworkopen.2024.30700
17. You J, Guo Y, Zhang Y, Kang JJ, Wang LB, Feng JF, Cheng W, Yu JT. Plasma proteomic profiles predict individual future health risk. *Nat Commun*. 2023;14:7817. doi: 10.1038/s41467-023-43575-7
18. Nurmohamed NS, Belo Pereira JP, Hoogeveen RM, Kroon J, Kraaijenhof JM, Waissi F, Timmerman N, Bom MJ, Hoefer IE, Knaapen P, et al. Targeted proteomics improves cardiovascular risk prediction in secondary prevention. *Eur Heart J*. 2022;43:1569–1577. doi: 10.1093/eurheartj/ehac055
19. 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
20. Matsushita K, Sang Y, Ning H, Ballew SH, Chow EK, Grams ME, Selvin E, Allison M, Criqui M, Coresh J, et al. Lifetime risk of lower-extremity peripheral artery disease defined by ankle-brachial index in the United States. *J Am Heart Assoc*. 2019;8:e012177. doi: 10.1161/JAHA.119.012177
21. Yuan S, Titova OE, Zhang K, Chen J, Li X, Klarin D, Åkesson A, Damrauer SM, Larsson SC. Circulating proteins and peripheral artery disease risk: observational and Mendelian randomization analyses. *Eur Heart J Open*. 2023;3:oead056. doi: 10.1093/ehjopen/oead056
22. Samman Tahhan A, Hayek SS, Sandesara P, Hajjari J, Hammadah M, O'Neal WT, Kelli HM, Alkhoder A, Ghasemzadeh N, Ko YA, et al. Circulating soluble urokinase plasminogen activator receptor levels and peripheral arterial disease outcomes. *Atherosclerosis*. 2017;264:108–114. doi: 10.1016/j.atherosclerosis.2017.06.019
23. Khalyfa A, Kheirandish-Gozal L, Bhattacharjee R, Khalyfa AA, Gozal D. Circulating microRNAs as potential biomarkers of endothelial dysfunction in obese children. *Chest*. 2016;149:786–800. doi: 10.1378/chest.15-0799
24. Zhao YY, Chen LH, Huang L, Li YZ, Yang C, Zhu Y, Qu SL, Zhang C. Cardiovascular protective effects of GLP-1: a focus on the MAPK signaling pathway. *Biochem Cell Biol Biochim Biol Cell*. 2022;100:9–16. doi: 10.1139/bcb-2021-0365
25. Fu J, Yu MG, Li Q, Park K, King GL. Insulin's actions on vascular tissues: physiological effects and pathophysiological contributions to vascular complications of diabetes. *Mol Metab*. 2021;52:101236. doi: 10.1016/j.molmet.2021.101236
26. Yamamoto M, Hanatani S, Araki S, Izumiya Y, Yamada T, Nakanishi N, Ishida T, Yamamura S, Kimura Y, Arima Y, et al. HE4 predicts progressive fibrosis and cardiovascular events in patients with dilated cardiomyopathy. *J Am Heart Assoc*. 2021;10:e021069. doi: 10.1161/JAHA.120.021069
27. Wang L, Hu N, Jiang Z, Luo P, Yi N, Chen Y, Zhou N, Liu L. Association between serum human epididymis protein 4 levels and cardiovascular events in obese patients with breast cancer. *Diab Metab Syndr Obes*. 2021;14:3703–3710. doi: 10.2147/DMSO.S320122
28. Lind L, Gigante B, Borné Y, Feldreich T, Leppert J, Hedberg P, Östgren CJ, Nyström FH, Sundström J, Ärnlöv J, et al. Plasma protein profile of carotid artery atherosclerosis and atherosclerotic outcomes: meta-analyses and Mendelian randomization analyses. *Arterioscler Thromb Vasc Biol*. 2021;41:1777–1788. doi: 10.1161/ATVBAHA.120.315597
29. Lind L, Gigante B, Borne Y, Mälarstig A, Sundström J, Ärnlöv J, Ingelsson E, Baldassarre D, Tremoli E, Veglia F, et al. The plasma protein profile and cardiovascular risk differ between intima-media thickness of the common carotid artery and the bulb: a meta-analysis and a longitudinal evaluation. *Atherosclerosis*. 2020;295:25–30. doi: 10.1016/j.atherosclerosis.2020.01.011
30. Goncalves I, Bengtsson E, Colhoun HM, Shore AC, Palombo C, Natali A, Edsfeldt A, Dunér P, Fredrikson GN, Björkbacka H, et al. Elevated plasma levels of MMP-12 are associated with atherosclerotic burden and symptomatic cardiovascular disease in subjects with type 2 diabetes. *Arterioscler Thromb Vasc Biol*. 2015;35:1723–1731. doi: 10.1161/ATVBAHA.115.305631
31. Mahdessian H, Perisic Matic L, Lengquist M, Gertow K, Sennblad B, Baldassarre D, Veglia F, Humphries SE, Rauramaa R, de Faire U, et al. Integrative studies implicate matrix metalloproteinase-12 as a culprit gene for large-artery atherosclerotic stroke. *J Intern Med*. 2017;282:429–444. doi: 10.1111/joim.12655
32. Li B, Shaikh F, Zamzam A, Syed MH, Abdin R, Qadura M. A machine learning algorithm for peripheral artery disease prognosis using biomarker data. *iScience*. 2024;27:109081. doi: 10.1016/j.isci.2024.109081
33. Chuang WC, Chu CH, Yao CS, Wei MC, Hsieh IL, Liao CM. The value of growth differentiation factor 15 as a biomarker for peripheral artery disease in diabetes patients. *Diabetol Metab Syndr*. 2025;17:31. doi: 10.1186/s13098-025-01588-w
34. Parvar SL, Fitridge R, Dawson J, Nicholls SJ. Medical and lifestyle management of peripheral arterial disease. *J Vasc Surg*. 2018;68:1595–1606. doi: 10.1016/j.jvs.2018.07.027
35. McCarthy CP, Shrestha S, Ibrahim N, van Kimmenade RRJ, Gaggin HK, Mukai R, Magaret C, Barnes G, Rhyne R, Garasic JM, et al. Performance of a clinical/proteomic panel to predict obstructive peripheral artery disease in patients with and without diabetes mellitus. *Open Heart*. 2019;6:e000955. doi: 10.1136/openhrt-2018-000955
36. Guo Y, You J, Zhang Y, Liu WS, Huang YY, Zhang YR, Zhang W, Dong Q, Feng JF, Cheng W, et al. Plasma proteomic profiles predict future dementia in healthy adults. *Nat Aging*. 2024;4:247–260. doi: 10.1038/s43587-023-00565-0
37. Xie R, Vlaski T, Sha S, Brenner H, Schöttker B. Sex-specific proteomic signatures improve cardiovascular risk prediction for the general population without cardiovascular disease or diabetes. *J Adv Res*. 2026;79:599–606. doi: 10.1016/j.jare.2025.03.034