Circulation: Genomic and Precision Medicine

ORIGINAL ARTICLE

Plasma Protein Profiling of Incident Cardiovascular Diseases: A Multisample Evaluation

Lars Lind, MD, PhD; Olga Titova, PhD; Rui Zeng, PhD; Daniela Zanetti, PhD; Martin Ingelsson, MD, PhD; Stefan Gustafsson, PhD; Johan Sundström, MD, PhD; Johan Ärnlöv, MD, PhD; Sölve Elmståhl, MD, PhD; Themistocles Assimes, MD, PhD; Karl Michaëlsson, MD, PhD

BACKGROUND: Proteomic profiling could potentially disclose new pathophysiological pathways for cardiovascular diseases (CVD) and improve prediction at the individual level. We therefore aimed to study the plasma protein profile associated with the incidence of different CVDs.

METHODS: Plasma levels of 245 proteins suspected to be linked to CVD or metabolism were measured in 4 Swedish prospective population-based cohorts (SIMPLER [Swedish Infrastructure for Medical Population-Based Life-Course and Environmental Research], ULSAM (Uppsala Longitudinal Study of Adult Men), EpiHealth, and POEM [Prospective Investigation of Obesity, Energy Production, and Metabolism]) comprising 11 869 individuals, free of CVD diagnoses at baseline. Our primary CVD outcome was defined by a combined end point that included either incident myocardial infarction, stroke, or heart failure.

RESULTS: Using a discovery/validation approach, 42 proteins were associated with our primary composite end point occurring in 1163 subjects. In separate meta-analyses for each of the 3 CVD outcomes, 49 proteins were related to myocardial infarction, 34 to ischemic stroke, and 109 to heart failure. Thirteen proteins were related to all 3 outcomes. Of those, urokinase plasminogen activator surface receptor, adrenomedullin, and KIM-1 (kidney injury molecule 1) were also related to several markers of subclinical CVD in Prospective Investigation of Obesity, Energy production and Metabolism, reflecting myocardial or arterial pathologies. In prediction analysis, a lasso selection of 11 proteins in ULSAM improved the discrimination of CVD by 3.3% ($P<0.0001$) in SIMPLER when added to traditional risk factors.

CONCLUSIONS: Protein profiling in multiple samples disclosed several new proteins to be associated with subsequent myocardial infarction, stroke, and heart failure, suggesting common pathophysiological pathways for these diseases. KIM-1, urokinase plasminogen activator surface receptor, and adrenomedullin were novel early markers of CVD. A selection of 11 proteins improved the discrimination of CVD.

Key Words: biomarkers ■ cardiovascular diseases ■ heart failure ■ ischemic stroke ■ myocardial infarction

Discovering biomarkers associated with incident disease could serve different purposes. First, those biomarkers could reflect underlying pathophysiological pathways that could be suitable targets for new therapies. Second, biomarker levels could be informative on the future risk of disease, thereby being suitable to be used for risk prediction purposes.

A large number of genetic susceptibility loci for 3 major cardiovascular diseases (CVDs) have been identified over the last 15 years including myocardial infarction, stroke, and heart failure (HF).[1–3] Some loci are shared between these diseases[4] suggesting a shared set of pathophysiological pathways for all 3 diseases. If functional, genetic loci could influence the levels of the corresponding proteins, and recently, a number of genetic loci being associated with protein levels have been published.[5] Proteomics bridges the gap between genomic information and functional proteins and translates this information to

Correspondence to: Lars Lind, MD, PhD, Department of Medical Sciences, Uppsala University Hospital, 751, 85 Uppsala, Sweden. Email lars.lind@medsci.uu.se
Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/CIRCGEN.123.004233.
For Sources of Funding and Disclosures, see page 543.

Circulation: Genomic and Precision Medicine is available at www.ahajournals.org/journal/circgen

Nonstandard Abbreviations and Acronyms

AUC	area under the curve
CVD	cardiovascular diseases
eGFR	estimated glomerular filtration rate
FGF-23	fibroblast growth factor 23
Gal-9	galectin-9
GDF-15	growth/differentiation factor 15
HF	heart failure
IL-6	interleukin 6
KIM-1	kidney injury molecule 1
MMP-12	matrix metalloproteinase-12
NT-proBNP	N-terminal prohormone brain natriuretic peptide
OPG	osteoprotegerin
OPN	osteopontin
PAI-1	plasminogen activator inhibitor 1
POEM	Prospective Investigation of Obesity, Energy Production and Metabolism
PON-3	paraoxonase-3
ROC	receiver operating characteristics
SIMPLER	Swedish Infrastructure for Medical Population-based Life-course and Environmental Research
TFF-3	trefoil factor 3
TNF-R2	tumor necrosis factor receptor 2
tPA	tissue-type plasminogen activator
TR	transferrin receptor protein 1
TRAIL-R2	TNF-related apoptosis-inducing ligand receptor 2
ULSAM	Uppsala Longitudinal Study of Adult Men
U-PAR	urokinase plasminogen activator surface receptor
WFDC2	WAP 4-disulfide core domain protein 2

a better understanding of etiologic pathways. In addition, early detection at the nonsymptomatic stages of chronic late-onset diseases by use of the proteome is also a conceivable key to provide a better outcome for therapeutic intervention. Emerging technologies provide new possibilities to discover novel and clinically relevant protein biomarkers in large-scale prospective studies.

In an attempt to search for proteins of importance for CVDs, we have recently published a single-center proteomics study in which the plasma levels of 41 proteins were related to a combined CVD end point.[6] When analyzing myocardial infarction, stroke, and HF as separate outcomes, only 3 proteins were related to each of these CVDs. However, that study was hampered by a low power due to a small sample size, and by the time of the investigation, we had no ability to replicate the findings in a separate cohort.

To overcome those problems, we have now performed a similar analysis, but in this case in more than 10× the number of subjects from 3 different cohorts (SIMPLER [Swedish Infrastructure for Medical Population-based Life-course and Environmental Research], ULSAM [Uppsala Longitudinal Study of Adult Men], and EpiHealth) using a strict discovery/validation approach for the combined CVD end point.

Using meta-analysis of these 3 cohorts, we also searched for proteins related to all 3 main CVDs to find pathophysiological pathways being in common for CVDs. As secondary aims, we did furthermore investigated if proteins of interest increased the discrimination of CVD compared with traditional CVD risk factors, and used a fourth cohort (POEM [Prospective Investigation of Obesity, Energy Production and Metabolism]) to evaluate if the proteins of interest also were related to subclinical markers of CVD in middle-aged individuals. The hypotheses tested were that we could identify more proteins being related to all 3 CVDs than published previously and that the CVD-related proteins also will be related to different markers of subclinical CVD in middle-aged subjects. It was furthermore hypothesized that CVD-related proteins will improve discrimination of CVD compared with established CVD risk factors.

METHODS

The complete methods section is given in the Supplemental Methods.

The study was approved by the Ethical Committee of Uppsala University and by the Swedish Ethics Authority. Each participant in the 4 study samples (EpiHealth,[7] ULSAM,[8] SIMPLER,[9] and POEM[10]) gave their informed consent.

The data that support the findings of this study are available from the corresponding author upon reasonable request.

RESULTS

Basic characteristics and number of incident cases of CVDs in the different cohorts are given in Table 1. In Table S1, the basic characteristics in SIMPLER, EpiHealth, and ULSAM are given divided into 3 groups; individuals without protein measurements, individuals with protein measurements with prevalent cardiovascular disease (CVD, myocardial infarction, ischemic stroke, or HF) at baseline (being excluded from the analysis), and individuals with protein measurements without prevalent CVD at baseline (the subjects at risk in the investigation).

The mean values for subclinical CVD markers in POEM are shown in Table S2.

Discovery and Replication of Proteins Related With the Composite CVD Outcome

In the discovery phase in SIMPLER, 114 proteins were related to incident CVD with FDR <0.05 in the age and sex-adjusted models, 42 of which were validated in the

Table 1. Basic Characteristics and Incident Cases of CVDs in the 4 Samples

	SIMPLER	ULSAM	EpiHealth	POEM
n	8701	826	2296	502
Age	71.5 (6.4)	71.2 (0.6)	61.1 (8.4)	50 (0.1)
Female sex	53.6%	0%	50%	50%
Systolic blood pressure, mm Hg	140.4 (17.8)	147 (19)	134.6 (17.0)	125.6 (16.4)
LDL-cholesterol, mmol/L	3.3 (1.0)	3.9 (0.9)	3.9 (1.0)	3.4 (0.9)
HDL-cholesterol, mmol/L	1.6 (0.4)	1.3 (0.4)	1.5 (0.3)	1.3 (0.3)
BMI, kg/m^2	26.3 (4.0)	26.3 (3.4)	26.5 (3.8)	26.4 (4.2)
Diabetes	11.1%	11%	8.3%	2.8%
Smoking	36.6%	21%	6.7 y of smoking	9.8%
Incident myocardial infarction (n)	255	125	61	NA
Incident ischemic stroke (n)	312	135	39	NA
Incident heart failure (n)	415	155	22	NA
Incident CVD (n, and median follow-up)	772 (5.1 y)	283 (12.5 y)	108 (8.5 y)	NA

Means and SDs are given, or proportions in %. BMI indicates body mass index; CVD, cardiovascular disease; HDL, high-density lipoprotein; LDL, low-density lipoprotein; NA, not assessed; POEM, Prospective Investigation of Obesity, Energy Production, and Metabolism; SIMPLER, Swedish Infrastructure for Medical Population-Based Life-Course and Environmental Research; and ULSAM, Uppsala Longitudinal Study of Adult Men.

meta-analysis of ULSAM and EpiHealth using FDR<0.05 in the age and sex-adjusted models and *P*<0.05 in the multiple-adjusted models.

We evaluated in the preanalysis stage if addition of estimated glomerular filtration rate (eGFR) to the multiple model including age, sex, and 6 traditional risk factors would change the HRs or not. In fact, the addition of eGFR to the multiple model did only change the HRs by on average <1% (eg, for NT-proBNP [N-terminal prohormone brain natriuretic peptide]) from 1.52 to 1.51, for MMP-12 (matrix metalloproteinase-12) from 1.22 to 1.21, for IL-6 (interleukin-6) from 1.15 to 1.15, and for KIM-1 (kidney injury molecule 1) from 1.22 to 1.22, to mention some important proteins). Thus, it is obvious that the inclusion of the 6 traditional risk factors also covered the potential confounding effects of eGFR, so the further addition of eGFR would only have a negligible effect on the estimates.

Figure 1 and Table S3 show the top 10 ranked validated proteins in terms of effect sizes for incident CVD and included NT-proBNP, GDF-15 (growth/differentiation factor 15), TRAIL-R2 (TNF-related apoptosis-inducing ligand receptor 2), MMP-12, PON3 (paraoxonase, inverse relationship), IL-6, KIM-1, Gal-9 (galectin-9), TFF3 (trefoil factor 3), and U-PAR (urokinase plasminogen activator surface receptor).

Meta-Analysis of Proteins Associated With Specific CVD outcomes

A meta-analysis of the 3 cohorts identified 113 proteins related to any of the 3 CVDs when considering an FDR<0.05 in the age and sex-adjusted models and a *P*<0.05 in the multiple risk factor–adjusted models (Figure S1). A total of 49 out of these 113 proteins were related to myocardial infarction (Table S4, 34 to ischemic stroke (Table S5), and 109 to HF (Table S6), while 13 were related to all 3 outcomes including TNF-R2 (tumor necrosis factor receptor 2), KIM-1, adrenomedulin, OPN (osteopontin), IL-6, U-PAR, OPG (osteoprotegerin), TR (transferrin receptor protein 1), TRAIL-R2, GDF-15, MMP-12, and FGF-23 (fibroblast growth factor 23; Figure 2).

Proteins Related With Subclinical Markers of CVD as Outcomes

We found these 13 proteins to be related to a number of subclinical markers of CVD (Figure 3; Table S7). We found adrenomedullin, KIM-1, and U-PAR to be linked to the largest number of subclinical markers of CVD, reflecting relationships with both the heart and the vasculature. The subclinical markers of VO_2max, Aix, and the E/A-ratio were related to the largest number of 13 proteins.

Prediction Model

Based on the discovery and replication results of our primary composite CVD outcome, a lasso regression conducted with the 42 validated proteins selected a subset of 11 proteins in the ULSAM study. The area under the curve (AUC) for the receiver operating characteristics (ROC) was found to be higher for these 11 proteins (ROC-AUC, 0.6949) compared with the ROC-AUC for traditional risk factors (ROC-AUC, 0.6226; *P*=0.0022), and these proteins provided a >8% (ROC-AUC, 0.7104; *P*<0.001) incremental improvement in the ROC-AUC over traditional risk factors alone (Table 2). When the same 11 proteins were evaluated in the SIMPLER Study (Figure 4; Table 2), the ROC-AUC was similar for the proteins (ROC-AUC, 0.7754) and the risk factors (ROC-AUC, 0.7705), but the addition of the 11 proteins to the risk factors increased the ROC-AUC by 3.3% compared with that with risk factors alone (ROC-AUC, 0.8038; *P*<0.0001).

DISCUSSION

The present study found 42 plasma proteins to be robustly associated with incident CVD using a discovery/validation approach in different samples. When the 3 different CVDs included in this composite CVD end point were analyzed separately, we found 13 proteins to

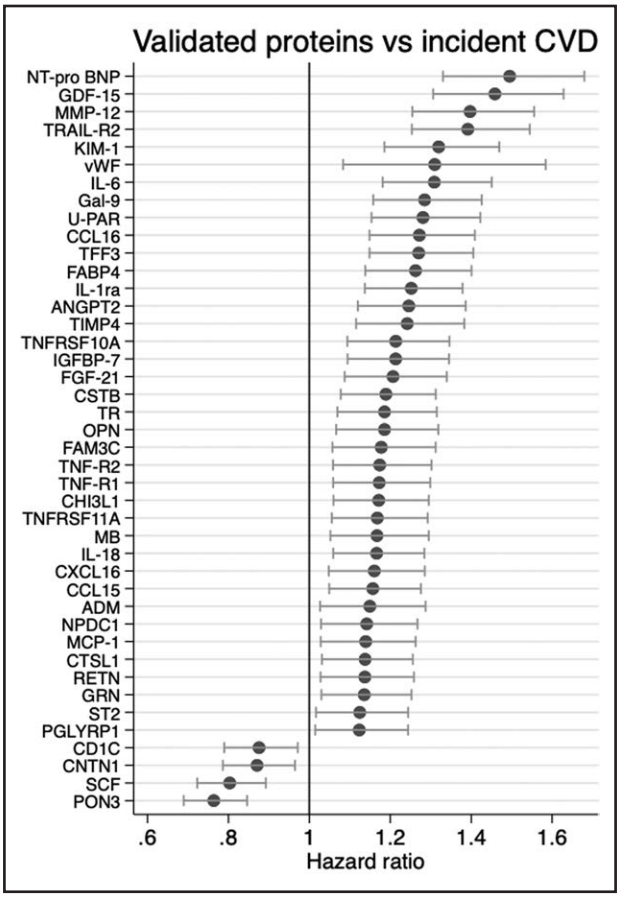

Figure 1. Overview of the 42 proteins that were validated to be associated with cardiovascular disease (CVD).
The estimates are from the meta-analysis of the ULSAM (Uppsala Longitudinal Study of Adult Men) and EpiHealth studies used in the validation step. Only the validated proteins are shown. ADM indicates proadrenomedullin; ANGPT2, angiopoietin 2; CCL16, C-C motif chemokine 16; CD1C, cluster of differentiation 1c, CHI3L1, chitinase-3-like protein 1; CNTN1, contactin-1; CSTB, cystatin-B; CTSL1, cathepsin L1; CXCL, C-X-C motif chemokine; FABP4, fatty acid–binding protein 4; FAM3C, family with sequence similarity 3c; FGF-21, fibroblast growth factor 21; Gal-9, galectin-9; GDF-15, growth/differentiation factor 15; GRN, granulins; IGFBP-7, insulin-like growth factor-binding protein 7; IL-1ra, interleukin-1 receptor antagonist protein; IL-6, interleukin-6; IL-18, interleukin-18; KIM-1, kidney injury molecule; MB, myoglobin; MCP-1, monocyte chemotactic protein 1; MMP-12, matrix metalloproteinase-12; NPDC1, neural proliferation differentiation and control protein 1; NT-proBNP, N-terminal prohormone brain natriuretic peptide; OPN, osteopontin; PGLYRP1, peptidoglycan recognition protein 1; PON3, paraoxonase; RETN, resistin; SCF, stem cell factor; ST2, interleukin 1 receptor-like 1; TFF3, trefoil factor 3; TIMP4, metalloproteinase inhibitor 4; TNF-R, tumor necrosis factor receptor; TNFRSF, tumor necrosis factor receptor superfamily member; TR, transferrin receptor protein 1; TRAIL-R2, TNF-related apoptosis-inducing ligand receptor 2; U-PAR, urokinase plasminogen activator surface receptor; and vWF, von Willebrand factor.

be related to all 3 outcomes. Of these 13 proteins, we found adrenomedullin, KIM-1, and U-PAR to be linked to the largest number of 12 subclinical markers of CVD measured in a fourth cohort. In prediction model analysis, a subset of 11 of the 42 replicated proteins from our primary analysis improved discrimination of incident CVD substantially when added on top of traditional risk factors.

Comparison With the Literature

Prior bidirectional 2-sample Mendelian randomization studies confirm the 3 major CVDs examined in this study possess a partial shared pathophysiology including a shared genetic susceptibility.[4] In this context, we asked to what degree are associations between plasma proteins and the same 3 CVDs shared. A prior study limited to 86 proteins found GDF-15 and TRAIL-R2 to be related to all 3 CVDs.[11] In another study, GDF-15, WFDC2 (WAP 4-disulfide core domain protein 2), and KIM-1 were found to be related to all 3 major CVD outcomes.[6] In the present study, we extend these efforts to multiple cohorts to substantially increase the power and found 13 proteins to be related to all 3 outcomes, of which 10 have not been previously described in this context.

Furthermore, a previous study restricted to a single cohort[6] found 41 out of 742 proteins measured in blood were related to the composite end point CVD compared with the present discovery/validation design, which found 42 of 245 investigated proteins to be associated with incident CVD. Importantly, only about a third of the proteins identified in the discovery phase could be validated, emphasizing the need for validation to reduce the number of false positive findings.

We also found that a subset of 11 proteins markedly improved discrimination of incident CVD when added on top of traditional risk factors building on our prior reports[6,11] and the large number of significant individual protein associations observed here. As expected, a larger improvement in discrimination was observed in the dataset in which the proteins were selected (ULSAM) compared with when the same set of proteins were evaluated in a separate sample (SIMPLER). The ROC-AUC for the traditional risk factors (including age and sex) was lower for ULSAM than for SIMPLER most likely because there was no variation in sex and minimal variation of age exists in ULSAM. Another recently published study using 27 selected proteins also reported an improvement in discrimination of CVD,[12] further emphasizing the concept that a signature of plasma protein has the potential to provide clinically meaningful improvement in risk prediction of CVDs.

A novel aspect of the present study was the ability to interrogate whether proteins related to all 3 CVDs were also related to markers of subclinical CVD in a middle-aged population free of CVD. We found adrenomedullin, KIM-1, and U-PAR to be linked to several of the markers of subclinical CVD, which further support the role of those proteins in early stages of pathological processes leading to clinically overt CVD. The relationship between these 3 markers and subclinical CVD was very

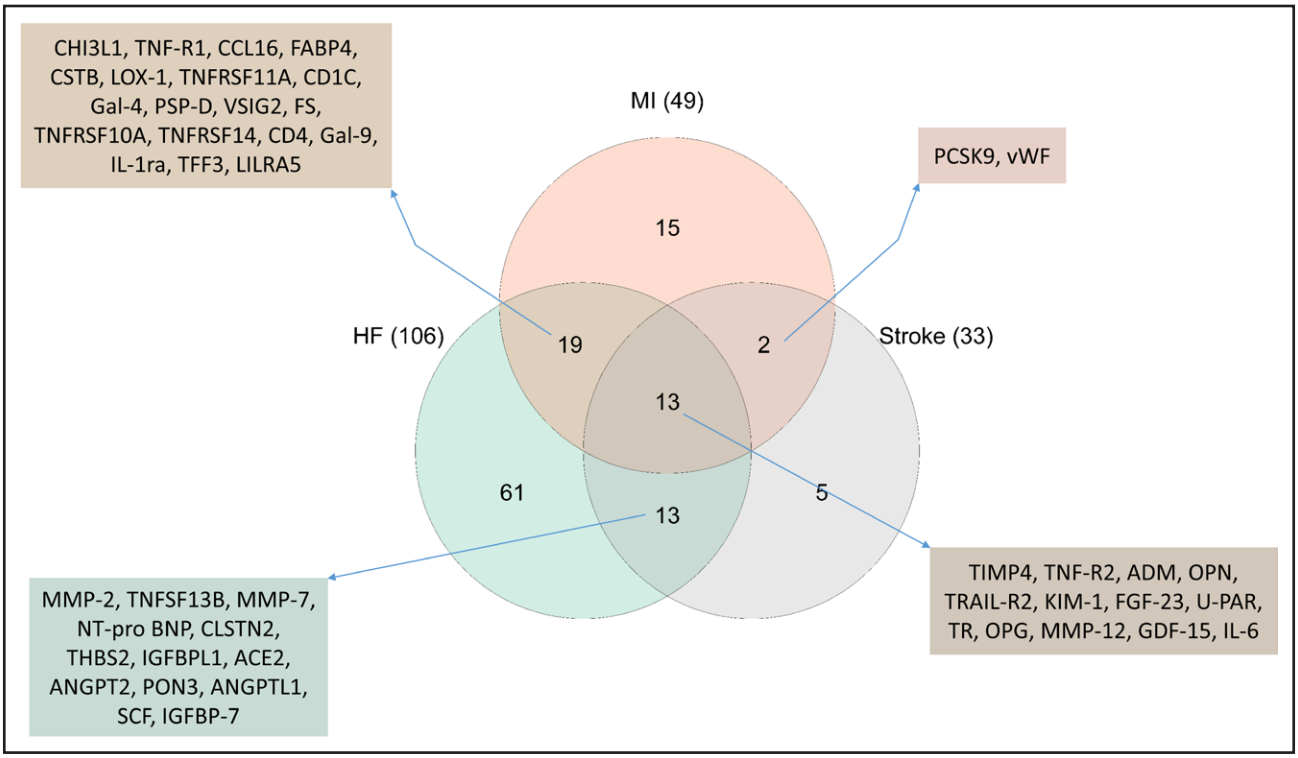

Figure 2. Venn diagram showing the overlap between the significant relationships for proteins vs myocardial infarction (MI), ischemic stroke and heart failure (HF).
Only the proteins being related to 2 or 3 of the cardiovascular diseases are shown. ACE2 indicates angiotensin-converting enzyme 2; ADM, proadrenomedullin; ANGPT2, angiopoietin 2; ANGPTL1, angiopoietin-related protein 1; CCL16, C-C motif chemokine 16; CD4, T-cell surface glycoprotein CD4; CD1C, cluster of differentiation 1c, CHI3L1, chitinase-3-like protein 1; CLSTN2, calsyntenin 2; CSTB, cystatin-B; FABP4, fatty acid–binding protein 4; FGF-23, fibroblast growth factor 23; FS, follistatin; Gal, galectin; GDF-15, growth/differentiation factor 15; IGFBP-7, insulin-like growth factor-binding protein 7; IL-1ra, interleukin-1 receptor antagonist protein; IL-6, interleukin-6; KIM-1, kidney injury molecule; LILRA5, leukocyte immunoglobulin-like receptor subfamily A member 5; LOX-1, lectin-like oxidized LDL receptor 1; MMP, matrix metalloproteinase; NT-proBNP, N-terminal prohormone brain natriuretic peptide; OPG, osteoprotegerin; OPN, osteopontin; PCSK9, proprotein convertase subtilisin/kexin type 9; PON3, paraoxonase; PSP-D, pulmonary surfactant-associated protein D; SCF, stem cell factor; TFF3, trefoil factor 3; THBS2, thrombospondin-2; TIMP4, metalloproteinase inhibitor 4; TNF-R, tumor necrosis factor receptor; TNFRSF, tumor necrosis factor receptor superfamily member; TNFSF13B, tumor necrosis factor ligand superfamily member 13B; TR, transferrin receptor protein 1; TRAIL-R2, TNF-related apoptosis-inducing ligand receptor 2; U-PAR, urokinase plasminogen activator surface receptor; VSIG2, V-set and immunoglobulin domain-containing protein 2; and vWF, von Willebrand factor.

similar to each other. Below, we provide further review and additional insights into each of these 3 proteins.

Adrenomedullin is a vasodilatory hormone expressed in a variety of tissues involved in angiogenesis, the function of the endothelial barrier, and the tolerance of the cell to oxidative stress. As reviewed by others,[13] circulating levels of adrenomedullin (or its propeptide) have been linked to HF and coronary heart disease but provide relatively less predictive ability compared with other biomarkers such as NT-proBNP and troponins. In the present study, adrenomedullin was inversely related to several of the vascular indices in the POEM Study, such as arm blood flow increase during hyperemia, distensibility, and echogenicity of the carotid artery, decreased diastolic function of the left ventricle (E/A-ratio and left atrial diameter [LA]), as well as poor exercise capacity. Whether or not adrenomedullin is causally involved in the development of these markers of early CVD, or rather is a compensatory mechanism, such as increased levels of NT-proBNP, remains to be established.

KIM-1 (kidney injury molecule 1, also known HAVcr-1 or TIM-1) is a transmembrane glycoprotein found on immune and epithelial cells. The protein is expressed in many tissues with the highest expression observed in kidneys. KIM-1 has an important role in the immune system but is also involved in the adaptive response following an acute kidney injury. Elevated plasma levels of KIM-1 have also been linked to CVDs and are also a powerful biomarker regarding risk of kidney failure in CVD patients.[14]

U-PAR is a part of the plasminogen activation system, which includes PAI-1 (plasminogen activator inhibitor 1) and tPA (tissue-type plasminogen activator). Apart from its proteolytic actions, U-PAR is also involved in cell migration, cell cycle regulation, and cell adhesion, and is a marker of systemic chronic inflammation.[15]

While adrenomedullin, KIM-1, and U-PAR were found to be related to not only overt CVD but also markers of

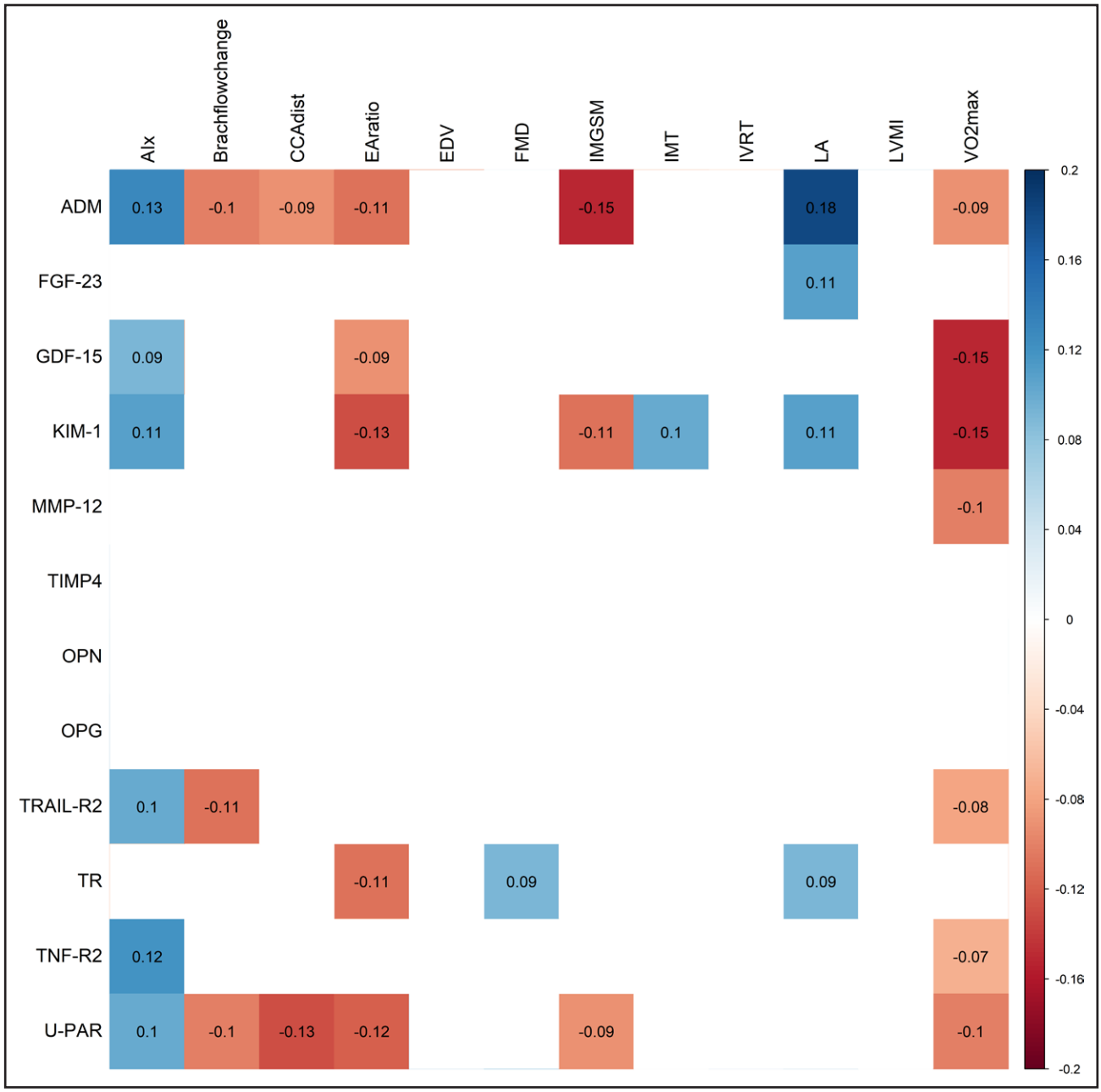

Figure 3. Relationships between the 13 proteins being related to all 3 cardiovascular diseases (CVDs) and subclinical markers of CVD in the POEM (Prospective Investigation of Obesity, Energy Production, and Metabolism) Study.
Regression coefficients are shown only for the relationships being significant. ADM indicates proadrenomedullin; Aix, augmentation index at radial artery pulse wave analysis; BrachFlowChange, increase in brachial artery blood flow following 5 minutes of blood flow arrest; CCAdist, carotid artery distensibility by ultrasound; EAratio, left ventricular diastolic function index; EDV, acetylcholine-mediated increase in forearm blood flow; FGF-23, fibroblast growth factor 23; FMD, brachial artery flow-mediated vasodilation; GDF-15, growth/differentiation factor 15; IMGSM, echolucency of the carotid artery intima-media complex; IMT, intima-media thickness of the carotid artery; LA, left atrial diameter; LVEDD, left ventricular end-diastolic diameter; LVMI, left ventricular mass; KIM-1, kidney injury molecule; MMP-12, matrix metalloproteinase-12; OPG, osteoprotegerin; OPN, osteopontin; TIMP4, metalloproteinase inhibitor 4; TNF-R2, tumor necrosis factor receptor 2; TR, transferrin receptor protein 1; TRAIL-R2, TNF-related apoptosis-inducing ligand receptor 2; U-PAR, urokinase plasminogen activator surface receptor; and VO2max, maximal oxygen consumption at an exercise test.

early and subclinical CVD, the same proteins could not be causally linked to coronary heart disease or stroke in a recent Mendelian randomization study.[5] Thus, further study is needed to be understand the role of these 3 studies in early alterations of the cardiovascular system, such as those studied in POEM.

The GFR in the kidneys is a major determinant of protein levels, and we have recently shown that the change in

Table 2. AUC for the ROC for the Comparison of Models With Risk Factors Only, 11 Selected Proteins Only, and Risk Factors Plus the Proteins

	ROC-AUC (95% CI)	Comparison
Discovery in ULSAM		
Risk factors only	0.6226 (0.58522–0.66006)	
11 proteins only	0.6949 (0.65957–0.73027)	Risk factors only vs proteins only: *P*=0.0022
Risk factor+proteins	0.7104 (0.67582–0.74508)	Risk factors only vs risk factors+proteins: *P*<0.0001
Validation in SIMPLER		
Risk factors only	0.7705 (0.75689–0.78409)	
11 proteins only	0.7754 (0.76196–0.78886)	Risk factors only vs proteins only: *P*=0.44
Risk factor+proteins	0.8038 (0.79103–0.81657)	Risk factors only vs risk factors+proteins: *P*<0.0001

The 11 proteins were selected by lasso regression in ULSAM and validated in SIMPLER. AUC indicates area under the curve; ROC, receiver operating characteristics; SIMPLER, Swedish Infrastructure for Medical Population-Based Life-Course and Environmental Research; and ULSAM, Uppsala Longitudinal Study of Adult Men.

the decline in GFR is a major determinant of the increase in protein levels generally seen over a 10-year period.[16] Since chronic renal failure also is a risk factor for CVD, we evaluated before we decided upon the variables to be included in the multiple model of the present investigation if the addition of eGFR to the 6 traditional CVD risk factors would impact the effect estimates. We then found that the addition of eGFR only had a negligible effect on the estimates. We therefore did not think it was of value to add eGFR to the multiple model, since it will only reduce the degree of freedom and thereby increase the *P* values. Thus, it is obvious that the inclusion of the 6 traditional risk factors also covered the potential confounding effects of eGFR.

HF is a diagnosis mainly based on the clinical presentation of the patient, but additional tests, such as NT-proBNP measurements are used to guide the diagnosis. It is therefore not strange that this protein is among the top findings for HF, representing a confirmation bias.

We have previously used the ULSAM Study to investigate relationships between proteins and CVD. In our first publication, the proteins were measured at age 77,[11] while in the second publication[6] and the present study, the proteins were measured at the age of 70 years. In the present study, the ULSAM sample (n=826 at risk of future CVD) was only 7% of the total number of individuals used for the evaluation of the associations (n=11 823) between proteins and incident CVD. Furthermore, the ULSAM data were used in the replication step together with another cohort (EpiHealth). Thus, even if ULSAM data play an important role also in the present study, the addition of >10 000 other individuals to the present investigation would ensure that we have the ability to find new associations not described in our previous publications.

The main strength of the present investigation is a combination of the large sample size involving multiple cohorts with the measure of a large number of proteins using the same measurements technique. This design allowed us to perform a powerful meta-analysis, to validate markers in independent datasets, and to explore relationships of markers of interest with subclinical markers of CVD. A major limitation is that the study participants are almost exclusively of European descent. Other race/ethnic groups must be studied to determine whether our findings are generalizable to other populations.

In conclusion, protein profiling in multiple samples disclosed several new proteins to be associated with myocardial infarction, stroke, and HF, suggesting common pathophysiological pathways for these diseases as has been observed in genetic studies. A selection of 11 such proteins substantially increased discrimination of CVD. Of these, KIM-1, U-PAR, and adrenomedullin were related to multiple markers of early CVD, as well as to overt CVD and deserve further study in additional cohorts with larger and more diverse participants.

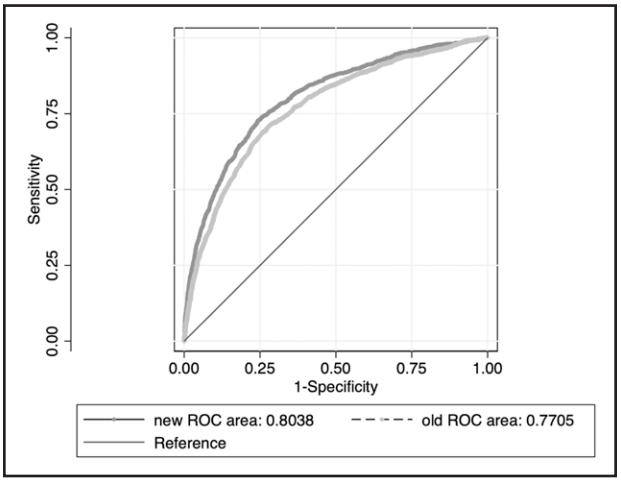

Figure 4. Area under the curve (AUC) for the receiver operating characteristics (ROC) for the comparison of models with risk factors only and risk factors plus the proteins in the SIMPLER (Swedish Infrastructure for Medical Population-Based Life-Course and Environmental Research) Study.
Old ROC area refers to the model with risk factors only. New ROC area refers to the model with risk factors plus the 11 proteins.

ARTICLE INFORMATION

Received March 7, 2023; accepted September 15, 2023.

Affiliations

Department of Medical Sciences (L.L., R.Z., S.G., J.S.), Department of Surgical Sciences (O.T., K.M.), and Department of Public Health and Caring Sciences/Geriatrics (M.I.), Uppsala University, Sweden. Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, CA (T.A., D.Z.). Division of Family Medicine and Primary Care, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Huddinge (J.A.). School of Health and Social

Sciences, Dalarna University, Falun (J.A.). Department of Clinical Sciences in Malmö, Lund University, Sweden (S.E.). Palo Alto VA Healthcare System, CA (T.A.).

Sources of Funding

This work was supported by a grant from the National Institute of Diabetes and Digestive and Kidney Diseases (1R01DK114183, PI: Assimes), part of the National Institutes of Health in the United States, and the Uppsala University Hospital (ALF-medel). SIMPLER (Swedish Infrastructure for Medical Population-based Life-course and Environmental Research) was supported by funding from the Swedish Research Council (https://www.vr.se; grants No. 2015 to 03257, 2017 to 00644, 2017 to 06100, and 2019 to 01291 to Karl Michaëlsson) and funding from Olle Engkvist Byggmästares stiftelse (SOEB). We acknowledge the national research infrastructure SIMPLER for generation and availability of data and computational facilities and resources. SIMPLER receives funding through the Swedish Research Council under the grant No. 2017 to 00644 and 2021 to 00160 (to Uppsala University and Dr Michaëlsson). The computations were performed on resources provided by the Swedish National Infrastructure for Computing's (http://www.snic.se) support for sensitive data Swedish National Infrastructure for Computing through the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX). Swedish National Infrastructure for Computing is financially supported by the Swedish Research Council. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the article.

Disclosures

Dr Gustafsson is employed by Sence Research AB. Dr Ärnlöv is a consultant for Boehringer Ingelheim and has financial relationships with Novartis and AstraZeneca. The other authors report no conflicts.

Supplemental Material

Supplemental Methods
Tables S1–S7
Figure S1
References 17–23

REFERENCES

1. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, Rutten-Jacobs L, Giese AK, van der Laan SW, Gretarsdottir S, et al; AFGen Consortium. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet*. 2018;50:524–537. doi: 10.1038/s41588-018-0058-3
2. Shah S, Henry A, Roselli C, Lin H, Sveinbjornsson G, Fatemifar G, Hedman AK, Wilk JB, Morley MP, Chaffin MD, et al; Regeneron Genetics Center. Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat Commun*. 2020;11:163. doi: 10.1038/s41467-019-13690-5
3. van der Harst P, Verweij N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ Res*. 2018;122:433–443. doi: 10.1161/CIRCRESAHA.117.312086
4. Lind L, Sundström J, Ärnlöv J, Ingelsson M, Henry A, Lumbers RT, Lampa E. Life-time covariation of major cardiovascular diseases: a 40-year longitudinal study and genetic studies. *Circ Genom Precis Med*. 2021;14:e002963. doi: 10.1161/CIRCGEN.120.002963
5. Folkersen L, Gustafsson S, Wang Q, Hansen DH, Hedman AK, Schork A, Page K, Zhernakova DV, Wu Y, Peters J, et al. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat Metab*. 2020;2:1135–1148. doi: 10.1038/s42255-020-00287-2
6. Lind L, Zanetti D, Ingelsson M, Gustafsson S, Ärnlöv J, Assimes TL. Large-scale plasma protein profiling of incident myocardial infarction, ischemic stroke, and heart failure. *J Am Heart Assoc*. 2021;10:e023330. doi: 10.1161/JAHA.121.023330
7. Lind L, Elmståhl S, Bergman E, Englund M, Lindberg E, Michaëlsson K, Nilsson PM, Sundström JE. A large population-based cohort study for investigation of gene-lifestyle interactions in the pathogenesis of common diseases. *Eur J Epidemiol*. 2013;28:189–197. doi: 10.1007/s10654-013-9787-x
8. Ärnlöv J, Ingelsson E, Sundström J, Lind L. Impact of body mass index and the metabolic syndrome on the risk of cardiovascular disease and death in middle-aged men. *Circulation*. 2010;121:230–236. doi: 10.1161/CIRCULATIONAHA.109.887521
9. Kaluza J, Stackelberg O, Harris HR, Akesson A, Björck M, Wolk A. Mediterranean diet is associated with reduced risk of abdominal aortic aneurysm in smokers: results of two prospective cohort studies. *Eur J Vasc Endovasc Surg*. 2021;62:284–293. doi: 10.1016/j.ejvs.2021.04.017
10. Lind L, Strand R, Michaëlsson K, Ahlström H, Kullberg J. Voxel-wise study of cohort associations in whole-body MRI: application in metabolic syndrome and its components. *Radiology*. 2020;294:559–567. doi: 10.1148/radiol.2019191035
11. Lind L, Ärnlöv J, Sundström J. Plasma protein profile of incident myocardial infarction, ischemic stroke, and heart failure in 2 cohorts. *J Am Heart Assoc*. 2021;10:e017900. doi: 10.1161/JAHA.120.017900
12. Williams SA, Ostroff R, Hinterberg MA, Coresh J, Ballantyne CM, Matsushita K, Mueller CE, Walter J, Jonasson C, Holman RR, et al. A proteomic surrogate for cardiovascular outcomes that is sensitive to multiple mechanisms of change in risk. *Sci Transl Med*. 2022;14:eabj9625. doi: 10.1126/scitranslmed.abj9625
13. Kita T, Kitamura K. Translational studies of adrenomedullin and related peptides regarding cardiovascular diseases. *Hypertens Res*. 2022;45:389–400. doi: 10.1038/s41440-021-00806-y
14. Karmakova capital Te CAC, Sergeeva NS, Kanukoev capital Ka C, Alekseev BY, Kaprin capital AC. Kidney Injury Molecule 1 (KIM-1): a multifunctional glycoprotein and biological marker (review). *Sovrem Tekhnologii Med*. 2021;13:64–78. doi: 10.17691/stm2021.13.3.08
15. Rasmussen LJH, Petersen JEV, Eugen-Olsen J. Soluble Urokinase Plasminogen Activator Receptor (suPAR) as a biomarker of systemic chronic inflammation. *Front Immunol*. 2021;12:780641. doi: 10.3389/fimmu.2021.780641
16. Lind L, Sundström J, Larsson A, Lampa E, Ärnlöv J, Ingelsson E. Longitudinal effects of aging on plasma proteins levels in older adults - associations with kidney function and hemoglobin levels. *PLoS One*. 2019;14:e0212060. doi: 10.1371/journal.pone.0212060
17. Assarsson E, Lundberg M, Holmquist G, Björkesten J, Thorsen SB, Ekman D, Eriksson A, Rennel Dickens E, Ohlsson S, Edfeldt G, et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One*. 2014;9:e95192. doi: 10.1371/journal.pone.0095192
18. Lind L, Siegbahn A, Lindahl B, Stenemo M, Sundström J, Ärnlöv J. Discovery of new risk markers for ischemic stroke using a novel targeted proteomics chip. *Stroke*. 2015;46:3340–3347. doi: 10.1161/STROKEAHA.115.010829
19. Lind L, Ärnlöv J, Lindahl B, Siegbahn A, Sundström J, Ingelsson E. Use of a proximity extension assay proteomics chip to discover new biomarkers for human atherosclerosis. *Atherosclerosis*. 2015;242:205–210. doi: 10.1016/j.atherosclerosis.2015.07.023
20. Merlo J, Lindblad U, Pessah-Rasmussen H, Hedblad B, Rastam J, Isacsson SO, Janzon L, Rastam L. Comparison of different procedures to identify probable cases of myocardial infarction and stroke in two Swedish prospective cohort studies using local and national routine registers. *Eur J Epidemiol*. 2000;16:235–243. doi: 10.1023/a:1007634722658
21. Ingelsson E, Ärnlöv J, Sundström J, Lind L. The validity of a diagnosis of heart failure in a hospital discharge register. *Eur J Heart Fail*. 2005;7:787–791. doi: 10.1016/j.ejheart.2004.12.007
22. Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. *Circulation*. 2016;133:601–609. doi: 10.1161/CIRCULATIONAHA.115.017719
23. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol*. 2009;170:244–256. doi: 10.1093/aje/kwp107