Original investigation

Sex-Specific Associations of the Plasma-Proteome with incident Coronary Artery Disease

Vincent Q. Sier1,2, Ko Willems van Dijk2,3,4, Diana van Heemst5, Paul H.A. Quax1,2, J. Wouter Jukema6,7, Raymond Noordam3*, Margreet R. de Vries1,2,8†

*shared last
†corresponding author

1Department of Surgery, Leiden University Medical Center, Leiden, the Netherlands; v.q.sier@lumc.nl; p.h.a.quax@lumc.nl; m.r.de_vries@lumc.nl
2Einthoven Laboratory for Experimental Vascular Medicine, Leiden University Medical Center, Leiden, the Netherlands
3Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands; k.willems_van_dijk@lumc.nl
4Department of Internal Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, the Netherlands;
k.willems_van_dijk@lumc.nl
5Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands;
d.van_heemst@lumc.nl; r.noordam@lumc.nl
6Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands; j.w.jukema@lumc.nl
7Netherlands Heart Institute, Utrecht, the Netherlands
8Department of Surgery, Brigham & Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

Abstract

Background and Aims
The aetiology of coronary artery disease (CAD) is different for men and women, yet insights into underlying sex-specific biological and pathophysiological mechanisms are limited. We investigated the sex-specific associations of the plasma-proteome with incident CAD.

Methods
In 40,829 participants from UK Biobank free-of-CAD from baseline to 365 days thereafter (55% women, mean 56.9±8.1 years), we analysed associations between 2,922 plasma-proteins and CAD incidence. Baseline plasma samples (2006-2010), were analysed in relation to incident CAD over a median follow-up of 13.7 (IQR: 13.1,14.4) years. Combined and sex-specific analyses were performed using Cox-proportional hazard models, adjusting for considered confounders, and causal inference using Mendelian Randomisation (MR).

Results
Multivariable-adjusted Cox-proportional hazard models identified 1,138 proteins associated with incident CAD (false-discovery-rate-corrected p-value<0.05), of which 219 showed evidence for potential causality using MR. Overrepresentation analyses identified involvement of cytokine-cytokine receptor interactions (p<0.0001), matrix remodelling (p<0.0001), regulation of innate and adaptive immune cells (p<0.0001), and angiogenesis (p<0.0001) pathways associated with incident CAD. Sex-specific analyses revealed additional 412 female-exclusive and 37 male-exclusive proteins and distinct CAD-risk pathways were identified for women (e.g., innate immune response) and men (e.g., tube morphogenesis (angiogenesis)). Translation toward druggability on targets with causal evidence revealed sex-specific clinical drug candidates such as C1S (men) and FOXO1 (women).

Conclusions
Although the majority of proteins showed consistent associations with incident CAD in both sexes, multiple proteins and biological pathways were either more strongly associated with incident CAD in men or in women, potentially indicating sex-specific pathogenesis and opening new alleys for prevention and clinical strategies.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Visual abstract
1. Introduction

Despite the growing recognition of disparities between men and women in the development, manifestation, and epidemiology of coronary artery disease (CAD), a critical gap persists in our understanding of its underlying sex-specific biological mechanisms. Both men and women are susceptible to develop ischaemic heart disease upon exposure to traditional risk factors. However, sex-specific aetiologies have been observed, with factors such as smoking, diabetes, and stress disproportionately affecting CAD incidence in women. Clinically, females present significantly more frequently with myocardial ischaemia with nonobstructive coronary arteries (INOCA), whereas males present more commonly with classical obstructive CAD. Regarding the latter, differences are observed in plaque composition, vulnerability, and burden between men and women, pointing toward a respectively more atheromatous versus fibrous phenotype.

The plasma proteome reflects the dynamic biological state, which, prior to disease, may provide insight into the mechanisms of (sex-specific) CAD onset. The UK Biobank is a large-scale biomedical healthcare database comprising population-scale data of 2,922 unique plasma proteins in 54,219 individuals. Illustratively, using the UK Biobank data, a recent study demonstrated the significance of large-scale integration of plasma protein levels and genomics.

Oomics approaches, including genomics, transcriptomics, metabolomics, and proteomics, are powerful tools to unravel underlying biological mechanisms of disease, and may also provide insights into the sex-specific onset of CAD. A limited number of genomic studies has identified multiple sex-specific candidate genes as risk factors for CAD, within risk loci that encode for proteins that were primarily engaged in processes involved in lipid metabolism and vascular remodelling. Focusing on lesion development, human regulatory network
analyses using patient material have indicated the existence of sex-specific plaque sub-phenotypes \textsuperscript{13,23} and cell signatures.\textsuperscript{24,25} 

Translating experimental findings into the identification of druggable targets is vital to enable clinical application. With the advent of increasingly complex datasets, advanced computational tools like interaction network analyses and artificial intelligence approaches facilitate the translation of novel targets to the clinic.\textsuperscript{26,27} 

We hypothesised that distinct proteomic patterns and pathways are associated with CAD onset in men and women. Therefore, in this study, by integrating both multivariable-adjusted regressions in combination with Mendelian Randomisation (MR) approaches,\textsuperscript{28} we aimed to investigate sex-specific plasma proteomic associations with incident CAD and identify potentially causal and druggable protein targets.
2. Methods

2.1 Multivariable-adjusted regression

Study setting (UK Biobank)

The UK Biobank is a general population cohort, in which approximately 500,000 participants were prospectively followed after recruitment between 2006 and 2010 (more information available via: https://www.ukbiobank.ac.uk/). Participants were between 40 and 70 years of age at enrolment and were recruited from the general population. Recruitment took place via invitation letters to eligible adults registered to the National Health Services (NHS) and living within 25 miles from one of the assessment centres. Written informed consent was retrieved from all participants and ethical approval of the study was given by the North-West Multicentre Ethics Committee. The present study was accepted and completed under project 56340.

Study population

The UK Biobank Pharma Plasma Proteome (UKB-PPP) project is a collaboration between thirteen biopharmaceutical companies and the UK Biobank, with the purpose of studying blood protein biomarkers in relation to disease onset. Baseline plasma samples from 54,219 individuals were selected randomly (n=46,595), or by consortium member pre-selection (n=6,376) and COVID-19 repeat imaging study (n=1,268) protocols. A total of 20 participants were pre-selected by the consortium members and participated in the COVID-19 repeat study.

For the current study, we restricted the analysis to participants of the UK Biobank from European ancestry to minimise population stratification bias. Moreover, individuals with a history of CAD or a CAD incident within 365 days after baseline were excluded for all analyses to minimise influences caused by reverse causation, resulting in a total cohort of 40,829 participants.
Exposure

Proteomic profiling on the UKB-PPP samples was performed using the Olink Explore 3072, which processes 2,941 protein analytes, corresponding to 2,923 unique proteins from the Cardiometabolic, Cardiometabolic_II, Inflammation, Inflammation_II, Neurology, Neurology_II, Oncology, and Oncology_II panels. Further details on the Olink explore platform and assays are available in the summary publication of the UKB-PPP.16 The current analyses have been performed in accordance with the quality consideration as described in Sun et al., excluding one protein (GLIPR1), due to >80% of data failing quality control (99.4%).16

Cardiovascular disease outcomes

A UK Biobank algorithm was employed to collect participant data on incident CAD, using data from the general practitioner, linked hospital admissions, death registries, and self-report. As coded according to the International Classification of Diseases (ICD) standards, the study outcome of CAD was defined as angina pectoris (I20), myocardial infarction (I21, I22), and acute and chronic ischaemic heart disease (I24, I25), whichever came first. Participants were followed until the occurrence of the CAD event, death, loss to follow-up or the end of follow-up, whichever came first.

Other variables

Data on age, sex, body mass index (BMI), insulin use, smoking, menopausal status (for women only), and use of cholesterol lowering medication were collected at the assessment centre of the UK Biobank through touchscreen questionnaire and physical measures. For smoking status, we included previous and current smokers. The Townsend Deprivation Index (TDI) was included as a marker for socio-economic status, and reflects the neighbourhood of
the individual participant based on zip code. Non-insulin dependent type diabetes mellitus was obtained via linkage to health-related records.

**Multivariable-adjusted Cox-proportional hazard models**

All the statistics were performed in R version 3.6.1 statistical software (The R Foundation for Statistical Computing, Vienna, Austria). Characteristics of the study population were studied at baseline as means (with standard deviations), medians (with interquartile ranges; for non-normally distributed continuous variables), or proportions (for categorical variables only).

Cox-proportional hazard models were performed to investigate the associations between standardised plasma protein levels (in standard deviation increase in level) and incident CAD in participants free-of-CAD at baseline or within 365 days thereafter (Survival package in R). In these analyses, participants were observed until the conclusion of the follow-up period, loss to follow-up or mortality, whichever came first. Regression analyses were adjusted for baseline age, sex, BMI, TDI, baseline diagnosis of diabetes mellitus, insulin use, smoking, and use of cholesterol-lowering medication.

Analyses were performed for the total study population, as well as stratified for men and women. Additional female analyses were performed to adjust and additionally restrict for post-menopausal status. To provide evidence favouring heterogeneity by sex, we additionally included a multiplicative interaction term between the protein level and sex in the analyses on incident CAD, adjusted for the confounders. We corrected for multiple testing using the false-discovery-rate (FDR).

For the analysis of enriched biological pathways in our data, we used the integrated biological database Metascape (web-based, more information available via: http://metascape.org). In short, Metascape facilitates the translation of large gene sets to involvement in biological processes via computed overrepresentation. More specifically,
Metascape incorporates multiple ontologies for enrichment analysis and eliminates redundancies following hierarchical clustering of terms based on enrichment. The platform leverages >40 independent knowledgebases, among which the ontology resources Molecular Signatures Database (MSigDB), Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO) knowledgebase, Reactome Pathway Database, and WikiPathways.32-36

2.2 Mendelian randomisation

To provide evidence favouring possible causal associations, we performed two-sample MR analyses. For the single-nucleotide polymorphisms (SNP)-exposure associations, we used the data from the SNPs associated with the protein levels from UK Biobank.16 From these, we selected all independent lead SNPs associated with protein levels (both cis and trans SNPs) with a SNP-exposure p-value <1.7e-11, considering multiple testing for the number of proteins in the genome-wide association study. As SNP-outcome datasets, we used summary-level genome-wide association data from the Coronary ARtery DIsease Genome wide Replication and Meta-analysis plus the Coronary Artery Disease Genetics consortium (CARDIoGRAMplusC4D) (60,801 cases and 123,504 controls),37 UK Biobank (122,733 cases and 424,528 controls), and freeze 9 from the Finngen biobank (https://www.finngen.fi/en) (43,518 cases and 333,759 controls), resulting in a total of 227,052 cases and 881,791 controls. In each dataset, CAD was defined as angina pectoris (I20), myocardial infarction (I21, I22), and acute and chronic ischaemic heart disease (I24, I25). With the exception of the CARDIoGRAMplusC4D dataset which comprised 23% of non-European ancestry participants, all other datasets only used participants from European ancestry. The studies contributing data to these (meta-)analyses have been approved by the necessary local medical ethics committees and all participants contributing to these efforts provided written informed consent. For the purpose of the present study, we only used the summary-level data and no
individual-level data. Data were only available for all participants combined and not stratified by sex.

**Mendelian Randomisation meta-analysis**

The primary analysis for MR employed inverse-variance weighted (IVW) regression analysis, assuming the absence of invalid genetic instruments such as directional pleiotropy. All MR analyses utilised the R-based package “TwoSample MR”. The mean effect estimate was derived separately from each outcome database through fixed-effect IVW meta-analysis of Wald ratios (gene-outcome [log odds ratio] divided by gene-exposure associations) estimated for each instrumental variable. Results were expressed as odds ratios (ORs) for CAD risk. Under the assumption of met MR criteria, this odds ratio served as an estimate of the causal effect of the exposure on the outcome.

To assess potential violations of main MR assumptions stemming from directional pleiotropy, MR-Egger regression analyses and weighted-median estimator analyses were performed. The MR-Egger's intercept estimated the average pleiotropic effect across genetic variants, with a (significant) non-zero value indicating bias in the IVW estimate. The weighted-median estimator yielded a consistent valid estimate if at least half of the instrumental variables were valid. All analyses were performed separately for the individual GWAS summary datasets, and subsequently meta-analysed using fixed-effects with the R-based “rmeta” package.

**2.3 Druggability analyses**

To evaluate target druggability, we used the machine learning platform DrugnomeAI to calculate aggregated probability scores based on a combined clinically-approved (Tclin) and clinical-phase (Tier 1) drug candidate model. In short, DrugnomeAI is an adaptation of
mantis-ml, a machine-learning framework, based on a stochastic semi-supervised learning approach.\textsuperscript{42,43}

Essentially, DrugnomeAI employs a process in which the machine learning algorithm learns from both labelled and unlabelled data in a probabilistic manner. Multiple resources have been integrated in the assessment of target druggability within DrugnomeAI, among which databases on protein-protein, drug-gene, and chemical-gene interactions such as Pharos, Drug Gene Interaction Database (DGIdb), and Comparative Toxicogenomics Database (CTDbase).\textsuperscript{44-46}
3. Results

3.1 Primary analysis UK Biobank

3.1.1 Summary information cases and controls

From the UKB-PPP project cohort, a total of 40,829 individuals were included free-of-CAD at baseline and in the first year after inclusion, with a mean age of 56.9 (standard deviation 8.10) years, 55.3% women, and a mean BMI of 27.0 kg/m² (Table 1). Cholesterol-lowering and insulin medication use at recruitment were 14.6% and 0.9% respectively, with 7.3% of total participants being diagnosed with non-insulin dependent diabetes mellitus. Moreover, there was a mean TDI (socio-economic status score) of -1.52 (3.01) and 44.6% of the individuals were past and/or current smoker (Table 1).

<table>
<thead>
<tr>
<th>Table 1: Baseline characteristics of the study population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Female (%)</td>
</tr>
<tr>
<td>Male (%)</td>
</tr>
<tr>
<td>Age at recruitment (years)</td>
</tr>
<tr>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Body mass index</td>
</tr>
<tr>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Cholesterol-lowering medication at recruitment</td>
</tr>
<tr>
<td>Yes (%)</td>
</tr>
<tr>
<td>Insulin use at recruitment</td>
</tr>
<tr>
<td>Yes (%)</td>
</tr>
<tr>
<td>Smoking</td>
</tr>
<tr>
<td>Past &amp; current (%)</td>
</tr>
<tr>
<td>Townsend deprivation index (socio-economic)</td>
</tr>
<tr>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Non-insulin dependent diabetes mellitus</td>
</tr>
<tr>
<td>Yes (%)</td>
</tr>
</tbody>
</table>
3.1.2 Proteomic associations with CAD

After a median follow-up time of 13.7 (IQR: 13.1, 14.4) years, there were 3,155 incident CAD cases, of which 38.7% were women. In combined sex multivariate analysis, 1,138 out of 2,922 proteins were significantly associated with incident CAD after correction for multiple comparisons (FDR) and adjusting for considered confounding factors (Figure 1a, Supplementary Table 1). Separate sex-specific analyses revealed 412 female- and 37 male-specific CAD proteins (Figure 1b, Supplementary Figure 1, Supplementary Table 2, Supplementary Table 3). Adjustment nor restriction of women for post-menopausal status did substantively change the results (Supplementary Figure 2, Supplementary Table 4).

![Figure 1](https://example.com/figure1)

**Figure 1. Proteomic associations for incident CAD.** (a) Volcano plot representing the effect size of combined-sex plasma-proteomic profiles for CAD (FDR<0.05, abs(β) ≥ 0.15). Orange dots denote positive significant associations with incident CAD. The proteins were part of the Cardiometabolic, Cardiometabolic_II, Inflammation, Inflammation_II, Neurology, Neurology_II, Oncology, and Oncology_II panels of the Olink Explore 3072 platform (b) Venn diagram of proteins associated with CAD for the sex-combined (green), female-specific (red), and male-specific (blue) Cox-proportional hazard models. FDR: false discovery rate

For graphical reasons only, we filtered for effect sizes ≥0.15 across the combined, female-, and male-specific analyses, resulting in 158 proteins of interest associated with incident CAD. Considering this cut-off, this resulted in 57 combined, 67 female-specific, and 37 male-specific proteins for visualisation (Figure 2).
Figure 2. Circular heatmap for proteins associated with increased hazard for incident CAD for men and women. Average effect sizes per protein are represented by the heatmap layer women (red; inner) and men (blue; outer) respectively. General clustering was plotted in a counter-clockwise fashion around the heatmap, based on increasing effect sizes and computed sex differences.

3.1.3 Sex-specific proteomic pathways

Based on the complete CAD-incidence-associated plasma-proteome in women and men, we performed pathway analyses. This revealed sex-specific differences associated with incident CAD (Figure 3, Supplementary Table 5, Supplementary Table 6). Notable distinctions were observed between both sexes, including angiogenic (tube morphogenesis) and hormonal
response (Insulin-like growth factor response) components in men, and the innate immune response and programmed cell death regulation in women (Figure 3a, 3b). Moreover, we also found multiple shared pathways between both sexes, among which those corresponding to the matrisome, cytokine-cytokine receptor interactions, regulation of cell migration, neutrophil degranulation, and enzyme-linked receptor protein signalling.

Figure 3. Sex-specific proteome pathway analyses for CAD incidence. (a, b) Pathway analyses of respectively the female and male CAD-associated proteome. Multiple ontology sources were included to cluster all enriched terms into groups.

(1) MSigDB: Molecular Signatures Database; (2) KEGG: Kyoto Encyclopedia of Genes and Genomes; (3) GO: Gene Ontology knowledgebase; (4) Reactome; (5) WP: WikiPathways.
3.1.4 Sex-different protein interaction analyses

Based on interaction analyses (male-versus-female),\textsuperscript{47} we found effect sizes of 711 proteins to be significantly different between men and women (FDR<0.05; Figure 4a, Supplementary Table 7). Based on these significantly different effect sizes in men versus women, we performed pathway analyses and subsequent network auto-clustering, and identified fourteen independent pathway clusters (Figure 4b, Supplementary Table 8). In general, the fourteen clusters represented differences between the CAD-associated proteome of men and women pertaining to innate and adaptive immunity, matrix remodelling, growth factor responses, apoptosis modulation, and lipid management. The three largest clusters were “response innate immune”, “pathogenic infection virus”, and “cysteine endopeptidase regulation” (Figure 4b).

a  Sex-different protein interactions heatmap

![Sex-different protein interactions heatmap](image)

b  Sex-different protein interaction clusters

![Sex-different protein interaction clusters](image)
Figure 4. Sex-different protein interactions. (a) Sex-protein interaction heatmap of significantly different proteins associated with CAD onset between men and women (n=711) (b) Network auto-clustering based on significant interaction clusters for incident CAD. Nodes represent individual pathways corresponding to the higher order cluster.

3.2 MR meta-analysis

We performed two-sample MR analyses on 2,346 proteins for which there were genetic instruments, and observed evidence for possible positive causal associations for 219 proteins in the meta-analysis of the three individual GWAS, of which 66 proteins had a logodds per s.d. higher level than 0.15 (Figure 5a, Supplementary Table 9).

To identify the sex-specific proteins with the highest hazard for CAD incidence, we analysed the 219 proteins for occurrence within our fourteen sex-protein interaction clusters (Supplementary Table 10). In total, we identified 59 hits that pertained to varying groups of clusters, of which the ten with the highest combined hazard ratio are shown in Figure 5b.

When considering the highest absolute differences between both sexes, Spondin 1 (SPON1), Complement C1s (C1S) and Cathepsin H (CTSH) emerged as additional top hits for men, while PDZ And LIM Domain 7 (PDLIM7), Drebrin like (DBNL), and Ubiquitin specific peptidase (USP8) were prominent within the specific top hits for women. Further examination revealed sex-specific nuances of the data within the framework of the fourteen pathway clusters (Figure 4b, Supplementary Table 10).
Figure 5. Causal evidence for sex-specific target associations with CAD. (a) Volcano plot of genetic MR associations with CAD in the combined UK Biobank-FinnGen-CARDIoGRAMplusC4D cohort. Orange and purple dots denote significant positive and negative associations respectively (FDR<0.05, abs(β) ≥ 0.15). (b) Forest plot depicting the top 10 target protein hazard ratios resulting from the sex-protein interaction analyses, as causally validated by the MR meta-analysis data (black) and jointly sorted for women (red) and men (blue). (c, d) Forest plot depicting the top three protein hazard ratios for CAD, as sorted for male- (c) and female- specific (d) differential top hits. *Clusters according to the interaction-analyses: 1: response innate immune, 2: actin based organisation, 3: cysteine endopeptidase regulation, 4: negative regulation kinase, 5: pathogenic infection virus, 6: erk1 positive regulation, 7: apoptosis modulation signalling, 8: lymphocyte proliferation activation, 9: membrane trafficking vesicle, 10: tyrosine growth factor, 11: cargo clathrin endocytosis, 12: fatty acids lipoprotein, 13: chaperone maturation folding, 14: vegfa vegfr2 signalling.
3.3 Top target druggability

Based on the 219 significant targets found through the MR analyses, we identified the top 20 hits associated with CAD across both sexes for the combined average of female and male hazard ratios as shown in Table 2. To assess druggability of these targets, we used aggregated probability scores of our top 20 hits based on a clinically-approved and clinical-phase drug candidate model (Table 2, Supplementary Table 11).

From this integrated data model, five targets emerged with high druggability probability scores surpassing the threshold of 0.5. Notably, TNFRSF4, CSF1, FURIN, CD74, TNFSF13 demonstrated potential for therapeutic intervention. (Table 2). Further analyses were performed using the sex-specific targets as identified for men and women. Within the previously identified male-specific top hits, the complement system activation component C1S had the highest druggability score (0.8578), followed by CTSH (0.5118) (Table 3). In the context of the female top hits, the first target with high druggability was Forkhead box protein 1 (FOXO1, 0.6286) (Supplementary Table 11).
<table>
<thead>
<tr>
<th>Target</th>
<th>Gene group</th>
<th>HR (95% CI) Women</th>
<th>HR (95% CI) Men</th>
<th>Druggability probability score (0-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDA2R (Ectodysplasin A2 receptor)</td>
<td>Tumour necrosis factor receptor superfamily</td>
<td>1.22 (1.16-1.28)</td>
<td>1.33 (1.26-1.4)</td>
<td>0.0779</td>
</tr>
<tr>
<td>COL6A3 (Collagen type VI alpha 3 chain)</td>
<td>Collagens</td>
<td>1.19 (1.13-1.25)</td>
<td>1.22 (1.17-1.27)</td>
<td>0.4555</td>
</tr>
<tr>
<td>CD74 (HLA class II histocompatibility antigen gamma chain)</td>
<td>CD molecules</td>
<td>1.24 (1.17-1.3)</td>
<td>1.19 (1.14-1.25)</td>
<td>0.9073</td>
</tr>
<tr>
<td>TNFRSF4 (Tumour Necrosis Factor receptor superfamily member 4)</td>
<td>CD molecules</td>
<td>1.21 (1.14-1.28)</td>
<td>1.19 (1.14-1.24)</td>
<td>0.9962</td>
</tr>
<tr>
<td>EFNA4 (Ephrin A4)</td>
<td>Ephrins</td>
<td>1.17 (1.11-1.24)</td>
<td>1.21 (1.16-1.26)</td>
<td>0.3482</td>
</tr>
<tr>
<td>TNFRSF14 (Tumour Necrosis Factor receptor superfamily member 14)</td>
<td>CD molecules</td>
<td>1.18 (1.12-1.25)</td>
<td>1.14 (1.09-1.19)</td>
<td>0.3397</td>
</tr>
<tr>
<td>CSF1 (Colon stimulating factor 1)</td>
<td>n.a.</td>
<td>1.19 (1.13-1.25)</td>
<td>1.15 (1.1-1.21)</td>
<td>0.9891</td>
</tr>
<tr>
<td>RSPO3 (R-Spondin 3)</td>
<td>R-spondin family</td>
<td>1.14 (1.07-1.22)</td>
<td>1.18 (1.13-1.23)</td>
<td>0.0234</td>
</tr>
<tr>
<td>CXCL9 (C-X-C motif chemokine ligand 9)</td>
<td>Chemokine ligands</td>
<td>1.17 (1.12-1.23)</td>
<td>1.14 (1.09-1.19)</td>
<td>0.495</td>
</tr>
<tr>
<td>SEPTIN8 (Septin 8)</td>
<td>Septins</td>
<td>1.17 (1.11-1.24)</td>
<td>1.11 (1.06-1.17)</td>
<td>0.0139</td>
</tr>
<tr>
<td>FURIN (Furin)</td>
<td>Proroprotein convertase subtilisin/kexin family</td>
<td>1.13 (1.06-1.2)</td>
<td>1.14 (1.08-1.2)</td>
<td>0.9354</td>
</tr>
<tr>
<td>BST2 (Bone marrow stromal cell antigen 2)</td>
<td>CD molecules</td>
<td>1.13 (1.08-1.19)</td>
<td>1.11 (1.06-1.15)</td>
<td>0.108</td>
</tr>
<tr>
<td>LGALS3 (Galectin 3)</td>
<td>Receptor ligands; Galectins</td>
<td>1.12 (1.06-1.19)</td>
<td>1.10 (1.05-1.15)</td>
<td>0.4985</td>
</tr>
<tr>
<td>MSR1 (Macrophage scavenger receptor)</td>
<td>Scavenger receptors;</td>
<td>1.14 (1.07-1.2)</td>
<td>1.14 (1.09-1.2)</td>
<td>0.5351</td>
</tr>
<tr>
<td>MDK (Midkine)</td>
<td>n.a.</td>
<td>1.11 (1.06-1.16)</td>
<td>1.13 (1.08-1.18)</td>
<td>0.4048</td>
</tr>
<tr>
<td>IL18BP (Interleukin 18 binding protein)</td>
<td>Immunoglobulin like domain containing</td>
<td>1.15 (1.09-1.21)</td>
<td>1.14 (1.09-1.19)</td>
<td>0.4106</td>
</tr>
<tr>
<td>TNFSF13 (Tumour Necrosis Factor superfamily member 13)</td>
<td>Tumour necrosis factor superfamily;</td>
<td>1.12 (1.06-1.18)</td>
<td>1.13 (1.08-1.18)</td>
<td>0.6179</td>
</tr>
</tbody>
</table>
Table 2: Druggability of top 20 combined targets

<table>
<thead>
<tr>
<th>Target</th>
<th>Gene group</th>
<th>HR (95% CI) Women</th>
<th>HR (95% CI) Men</th>
<th>Druggability probability score (0-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCL21</td>
<td>C-C motif chemokine ligand 21</td>
<td>Chemokine ligands</td>
<td>1.22 (1.16-1.28)</td>
<td>1.15 (1.11-1.19)</td>
</tr>
<tr>
<td>BTN3A2</td>
<td>Butyrophilin subfamily 3 member A2</td>
<td>Butyrophilins V-set domain containing</td>
<td>1.19 (1.13-1.25)</td>
<td>1.11 (1.06-1.16)</td>
</tr>
<tr>
<td>LTBP3</td>
<td>Latent transforming growth factor beta binding protein 3</td>
<td>Latent transforming growth factor beta binding proteins</td>
<td>1.24 (1.17-1.3)</td>
<td>1.14 (1.09-1.18)</td>
</tr>
</tbody>
</table>

Table 3: Druggability of top three female- and male-specific proteins

<table>
<thead>
<tr>
<th>Target</th>
<th>Gene group</th>
<th>HR (95% CI) Women</th>
<th>HR (95% CI) Men</th>
<th>Druggability probability score (0-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Women</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PDLIM7</td>
<td>PDZ and LIM domain 7</td>
<td>PDZ domain containing LIM domain containing</td>
<td>1.01 (0.96-1.05)</td>
<td>1.15 (1.09-1.22)</td>
</tr>
<tr>
<td>DBNL</td>
<td>Drebrin like</td>
<td>MicroRNA protein coding host genes</td>
<td>0.96 (0.94-1.03)</td>
<td>1.13 (1.06-1.19)</td>
</tr>
<tr>
<td>USP8</td>
<td>Ubiquitin specific peptidase 8</td>
<td>Ubiquitin specific peptidases</td>
<td>0.96 (0.93-1.02)</td>
<td>1.11 (1.05-1.18)</td>
</tr>
<tr>
<td><strong>Men</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPON1</td>
<td>Spondin 1</td>
<td>n.a.</td>
<td>Sushi domain containing</td>
<td>1.17 (1.12-1.22)</td>
</tr>
<tr>
<td>C1S</td>
<td>Complement C1S</td>
<td>Complement system activation components</td>
<td>1.12 (1.07-1.18)</td>
<td>1.02 (0.95-1.08)</td>
</tr>
<tr>
<td>CTSH</td>
<td>Cathepsin H</td>
<td>Minor histocompatibility antigens</td>
<td>1.11 (1.06-1.16)</td>
<td>0.99 (0.94-1.05)</td>
</tr>
</tbody>
</table>
4. Discussion

This study observed the presence of sex-specific differences in the plasma proteome associated with incident CAD. Through integrated genomic and proteomic approaches, we observed differences between men and women in pathways related to matrix organisation, cytokine interactions, regulation of innate and adaptive immune cells, angiogenesis, and growth factor signalling related to incident CAD. Furthermore, we provided evidence favouring causal relations in a sex-combined study sample. These findings granted new insights into sex-specific pathophysiological mechanisms and implications for incident CAD, and allowed for the identification of druggable targets unique to men and women.

Despite apparent differences in top target findings of previous plasma proteomic approaches,\textsuperscript{16,17,48-50} we found evidence for the contribution of shared biological pathways leading to CAD. Notably, from our protein-sex interaction analyses it appeared that specific pathways, such as vascular endothelial growth factor (VEGF) signalling, are associated with sex differences in CAD development. Previously, our group and others have shown that dysfunctional VEGF signalling is involved in leaky plaque neovessels and can subsequently contribute to lesion destabilisation.\textsuperscript{51-53} Non-mature neovessels promote plaque development as the endothelium functions as a significant entrance route for immune cells via NF-κB promotion and subsequent TNF signalling.\textsuperscript{54,55} Endothelial dysfunction has been described in both sexes within the context of atherosclerotic disease. Men are more likely to suffer from intraplaque haemorrhage secondary to neovessel-rupture, whereas women more often present with coronary microvascular dysfunction (CMD) and subsequent angina with non-obstructive coronary arteries.\textsuperscript{56-58} This is noteworthy, considering that female hormone signalling has been associated with CMD,\textsuperscript{59} while oestrogen replacement therapy has not been found effective to halt classical obstructive atherosclerotic plaque progression in post-menopausal women with ischaemic heart disease.\textsuperscript{60,61} The findings of our present study further substantiate the concept
that in the development of CAD, both sexes exhibit numerous common risk factors in addition to sex-specific mechanisms. Specifically, our results highlight the involvement of various conventional protein risk factors such as the well-described targets LDLR and PCSK9, which are shared between women and men.

After integration of large-scale plasma proteomics with MR analyses, we found evidence of high druggability for 5/20 of our strongest CAD-associated hits (TNFRSF4, TNFSF13, CSF1, FURIN, CD74). Indeed, tumour necrosis factor levels have previously been associated with higher risk of coronary artery disease and specific drugs targeting TNFRSF4 and TNFSF13 have been employed in clinical trials in patients with respectively T- and B-cell related auto-immune diseases such as IGA nephropathy (Sibeprenlimab) and atopic dermatitis (telazorlimab).\textsuperscript{62-66} CSF1 is a known driver of atherosclerosis through monocyte/macrophage activation, proliferation, and differentiation.\textsuperscript{67} High levels of CSF1 expression in the tumour micro-environment have been related to poor prognosis in solid tumours, in which context numerous monoclonal antibody therapies have reached clinical trials with mixed results.\textsuperscript{68} Recently, FURIN, a member of the proprotein convertases, was found by Mazidi et al. as their most strongly associated protein with ischaemic heart disease\textsuperscript{17} and was previously linked to atherogenesis via upregulation of TNFSF13.\textsuperscript{69-71} Another hit, CD74, is expressed on monocytes, macrophages, and B-cells and is involved in antigen presentation. Direct inhibition of CD74 through the monoclonal antibody milatuzumab has been described in haematologic malignancies.\textsuperscript{72,73} Another interesting approach is to induce CD74 degradation through inhibition of the cysteine protease Cathepsin S. Elevated serum cathepsin S levels are associated with plaque instability and vulnerability and a small molecule inhibitor has been described for patients with Sjögren’s syndrome (petesicatib) (clinicaltrials.gov: nct02701985).

Taking sex-specificity into consideration, we identified 59 sex-specific associations with evidence for possible causality, of which C1S (complement factor) and FOXO1 (transcription
factor) were the top druggable hits in men and women respectively. From the literature, it is known that general C1 inhibition diminishes early intimal hyperplasia. Conversely, deficiencies in C1 inhibitory protein function are associated with hereditary angioedema and subsequent microvascular endothelial dysfunction within the atherosclerotic plaque, drawing parallels with the tube morphogenesis pathway in our male-specific analyses.

Although we have found many more proteins to be specifically associated with incident CAD in women compared to men, despite the lower number of cases in women, the strongest protein-CAD associations in women did not seem to yield strong evidence as druggable targets using currently available databases. The first druggable target in women was FOXO1, a transcription factor which has also been described to be involved in the pathogenesis of polycystic ovary syndrome and glucose metabolism. Altogether, our findings point toward sex-different pathways and factors in CAD development, and suggest sex-specific drug candidates among our identified hits.

Strengths & limitations

The primary strength of our study is the large sample size and in particular the proportion of included women, enabling the analysis of a comprehensive pool of plasma proteins in association with CAD while correcting for multiple testing and confounders. The integration of genomics and proteomics provided evidence for causality. Limitations include the focus on individuals of European ancestry, and therefore these results apply to only that part of the population. We did not have access to sex-stratified summary-level datasets of GWAS on CAD. For this reason, we were not able to perform MR analyses in men and women separately. Alternatively, we performed MR analyses in a large sample of men and women combined. Furthermore, in the multivariable-regression analyses, we excluded participants who developed CAD within the first year of follow-up to limit the potential influence of reverse
causation. Nevertheless, this approach does not limit the potential presence of sex-specific residual confounding, although adjustment nor restriction of women for postmenopausal status did materially differ the results, suggesting such confounding might be limited.

Conclusion

In summary, by linking the sex-specific plasma-proteome to CAD incidence and providing evidence for causal relations, we were able to identify strongly-associated female- and male-specific proteins along with their biological pathways. These findings offer potential directions for developing targeted preventive and interventional strategies against CAD.

Disclosure of interest

All authors declare no conflict of interest.

Data Availability

The present study has been conducted using the UK Biobank Resource (Application Number 56,340) that is available to researchers. Summary data are available from the original sources. Summary-level data used for the two-sample Mendelian Randomisation analyses are available online through the infrastructure of the R package twosampleMR or on the website of the FinnGen Biobank (https://www.finngen.fi/en)

Author contribution statement

Conceptualisation; VQS, JWJ, PHAQ, RN, MRdV. Methodology; VQS, RN. Formal analysis; VQS, RN. Resources; RN, KWvD, DvH. Writing- Original Draft; VQS, RN, MRdV. Writing- Review & Editing; VQS, KWvD, DvH, JWJ, PHAQ, RN, MRdV. Visualisation;
VQS. Supervision; PHAQ, RN, MRdV, Project Administration; RN, MRdV. Funding acquisition: VQS, PHAQ, RN, MRdV.

**Funding**

This work was supported by a Leiden University Medical Centre MD/PhD research grant to V.Q. Sier.
References


