This article has been accepted for publication in HEARTJNL following peer review.
The definitive copyedited, typeset version is available online at
10.1136/heartjnl-2024-324050

Associations between circulating proteins and cardiometabolic diseases: a systematic review and meta-analysis of observational and Mendelian randomization studies

Running title: Proteomics and cardiometabolic diseases

Ting Wu, MSc[1]*, Yalei Ke, MSc[1]*, Yingtao Li, MSc[1], Zhiyu Wu, MSc[1], Jun Lv, PhD[1,2,3], Canqing Yu, PhD[1,2,3], Dianjianyi Sun, PhD[1,2], Pang Yao, PhD[4,5], Christiana Kartsonaki, DPhil[4,5], Zhengming Chen, DPhil[4,5], Liming Li, MD[1,2,3], Yuanjie Pang, DPhil[1,3]

1. Department of Epidemiology & Biostatistics, School of Public Health, Peking University, Beijing 100191, China;
2. Peking University Centre for Public Health and Epidemic Preparedness & Response, Beijing 100191, China;
3. Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing, China
4. Medical Research Council Population Health Research Unit at the University of Oxford, Oxford, United Kingdom
5. Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford, United Kingdom

* Joint first author.

Address for correspondence:

Dr. Yuanjie Pang
Dept. of Epidemiology and Biostatistics
School of Public Health
Peking University
38 Xueyuan Road
Beijing, 100191, China
Tel: 86-010-82801528
Fax: 86-010-82801530
yuanjie_p@163.com

Word count: Abstract 278, Text 2999
6 figures, 31 eTables/eFigures

Revised in July 2024

Abstract

Background: Integration of large proteomics and genetic data in population-based studies can provide insights into discovery of novel biomarkers and potential therapeutic targets for cardiometabolic diseases (CMD). We aimed to synthesize existing evidence on the observational and genetic associations between circulating proteins and CMD.

Methods: PubMed, Embase, and Web of Science were searched until July 2023 for potentially relevant prospective observational and Mendelian randomization (MR) studies investigating associations between circulating proteins and CMD, including coronary heart disease, stroke, type 2 diabetes, heart failure, atrial fibrillation, and atherosclerosis. Two investigators independently extracted study characteristics using a standard form and pooled data using random effects models.

Results: 50 observational, 25 MR, and ten studies performing both analyses were included, involving 26,414,160 non-overlapping participants. Meta-analysis of observational studies revealed 560 proteins associated with CMD, of which 133 proteins were associated with ≥two CMDs (i.e. pleiotropic). There were 245 potentially causal protein biomarkers identified in MR pooled results, involving 23 pleiotropic proteins. IL6RA and MMP12 were each causally associated with seven diseases. 22 protein-disease pairs showed directionally concordant associations in observational and MR pooled estimates. Addition of protein biomarkers to traditional clinical models modestly improved the accuracy of predicting incident CMD, with the highest improvement for heart failure (ΔC-index ~0.2). Of the 245 potentially causal proteins (291 protein-disease pairs), three pairs were validated by evidence of drug development from existing drug databases, 288 pairs lacked evidence of drug development, and 66 proteins were drug targets approved for other indications.

Conclusions: Combined analyses of observational and genetic studies revealed the potential causal role of several proteins in the aetiology of CMD. Novel protein biomarkers are promising targets for drug development and risk stratification.

Keywords*: cardiometabolic disease; proteomics; Mendelian randomization; meta-analysis*

WHAT IS ALREADY KNOWN ON THIS TOPIC

- We searched PubMed, Web of Science, and Embase from database inception to July 2023, using search terms pertaining to cardiometabolic diseases (CMD) and proteomics without language restrictions. In the past decade, hundreds of large population-based observational and genetic studies have investigated the associations between circulating proteins and CMD. Because of the variation in study designs, sample sizes, and proteomic methodologies, the associations between circulating proteins and CMD have been inconclusive.

WHAT THIS STUDY ADDS

- The present study systematically assessed the direction and magnitude of the associations between circulating proteins and CMD. Meta-analyses of observational and MR studies identified 560 and 245 CMD-associated proteins, respectively. Out of 291 Tier 1 or 2 protein-disease pairs, 288 showed no evidence in drug development databases, and 66 proteins were recognized as drug targets approved for other indications. Furthermore, integration of protein biomarkers into traditional clinical models modestly enhanced the prediction of incident CMD.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- This systematic review and meta-analysis provides a thorough evaluation of the current evidence on the role of circulating proteins in CMD. By integrating proteomics and genomics, the approach we adopted can inform the selection of protein biomarkers to improve risk stratification of CMD. Additionally, this method can be utilized in the early stages of drug discovery to identify promising targets, and it can be integrated with traditional approaches to improve the assessment of drug repurposing opportunities.

Introduction

Cardiometabolic diseases (CMDs) are the leading cause of death and disability globally.[1] Proteins play critical roles in the biological processes involved in CMD and constitute effective drug targets.[2] Proteomics informs the holistic and comprehensive understanding of molecular and cellular mechanisms underlying the pathogenesis of diseases.[3] In recent years, high-throughput proteomics assays have flourished and proteomics has been widely used in large-scale population-based studies.

Mendelian randomization (MR) utilizes the random allocation of genetic alleles during meiosis and uses genetic variants specifically related to a particular exposure to examine the causal effect of the exposure on the disease.[4] Compared with observational studies, MR studies seek to establish whether specific proteins are causally related to CMD risk or represent downstream markers of CMD-related processes. By providing evidence of causation, MR has the potential to accelerate genetics-guided drug discovery.[5]

In the past decade, hundreds of observational and MR studies have examined the associations between circulating proteins and CMD. However, the study design, sample size, and proteomic assays varied across studies. Therefore, we undertook a systematic review and meta-analysis to assess the direction and magnitude of the associations between circulating proteins and CMD, so as to provide clues and references for research on potential biological mechanisms and drug targets of CMD.

Materials and Methods

Literature search, study selection and data extraction

We systematically searched PubMed, Embase, and Web of Science from inception to 11 July 2023. We included prospective observational and MR studies that investigated the associations between circulating proteins and CMD, including coronary heart disease (CHD, I20-I25), stroke (I60-I69), type 2 diabetes (T2D, E11), heart failure (HF, I50), atrial fibrillation (AF, I48), and atherosclerosis. When two or more studies reported data from the same cohort or consortium, only the study with the largest number of participants was included. We excluded (1) studies that used post-mortem blood, tissue or urine samples; (2) studies that recruited patients with non-CMD (e.g. dementia, arthritis) at baseline; and (3) *in vivo* and *in vitro* studies. We also manually searched the reference list of the retrieved review articles to identify other studies. No language restriction was imposed, and all included studies were in English. Studies were evaluated against the inclusion and exclusion criteria by two independent researchers and any difference were resolved via discussion with a third researcher. Data from included studies were extracted into predefined tables by two researchers independently. The study protocol was registered with PROSPERO (Registration number: CRD42022350327). Detailed search strategies and data extraction procedures were shown in eMethods.

Quality assessment

Quality assessment was conducted for observational and MR studies, separately. For observational studies, the quality of each study was assessed according to the Newcastle-Ottawa Scale (NOS) by two reviewers independently (T.W. and Y.K.). NOS covered three domains: subject representativeness, comparability of subjects, and ascertainment of risk. The length of follow-up was set at a minimum of five years to be considered as adequate. NOS scores were categorized as

high quality (seven to nine stars), moderate quality (four to six stars), and low quality (zero to three stars).

Statistical analysis

Before meta-analyses, we standardized names of proteins according to Unified Protein Database (Uniprot) (Table S1)[6]. Meta-analyses were performed separately for observational and MR studies, and for each specific disease outcome. Protein-disease pairs with ≥two non-overlapping reports were included. Relative risk (RR) estimates for observational studies and odds ratio (OR) estimates for MR studies per one standard deviation (SD) increase in proteins levels were pooled using random effects models via R package 'metafor'. Between-study heterogeneity was assessed using the I^2 statistic.

For MR studies, analyses were restricted to European population because 91.4% studies were conducted in Europeans, and excluded proteins using *trans*-protein quantitative trait loci (pQTL) to avoid horizontal pleiotropy. We graded the evidence of proteins in MR studies (Figure 1). Meanwhile, significant protein-disease pairs reported by only one MR study were also graded and presented for their valuable insights into causality. For proteins with Tier 1 or 2 MR evidence (i.e. top levels of certainty), we conducted pathway enrichment analysis[7] and evaluated the druggability[8] (eMethods).

To enhance the credibility and interpretability of our results, we compared findings from observational and MR studies. A protein-disease pair was considered consistent if: 1) the association was significant in observational meta-analysis and graded as Tier 1-2 on MR evidence, and 2) the directions of effect estimates were concordant in both analyses.

A 2-sided p-value <0.05 was considered statistically significant. A Benjamini-Hochberg false discovery rate (FDR) <5% was used to account for multiple comparisons. All statistical analyses were conducted using R V.4.2.2.

Patient and public involvement in research

We report no patient or public involvement in the design or implementation of the study.

Results

The overview of analytic approaches and key findings were presented in Figure 1. The literature search generated 14,932 records, and 85 studies were included in the final analysis, involving 50 prospective observational studies, 25 MR studies, and ten studies performing both observational and MR analyses (Figure S1). The characteristics of included studies were summarized in Table S2. For a full reference list, see Supplement 1.

Observational associations between proteins and CMD

A total of 60 studies examined the associations between proteins and incident CMDs, reporting results for 3788 protein-disease pairs. 2318 pairs with two or more reports were included in meta-analysis. Of these, the associations of 748 pairs remained significant in meta-analysis (Figure 2-3). The number of proteins included in each stage is summarized by diseases in Figure 2. Among all stroke subtypes, only incident ischemic stroke (IS) was investigated and included in meta-analysis. Moderate heterogeneity was observed for observational pooled results, and 45.8% pairs had $I^2 \geq 80\%$. Detailed effect estimates of meta-analysis specific for each disease were summarized in Table S3-S8.

In our pooled results, 133 proteins were associated with risk of two or more CMDs, referred to as "pleiotropic protein" (Figure 4). These included 94 proteins associated with two diseases, 27 proteins with three diseases, nine proteins (FABP4, IBP2, IL6, MMP12, ANFB, TNR1B, TR10B, UPAR, HGF) with four diseases, and three proteins (GDF15, HAVR1, MMP7) with five diseases. The directions and strengths of associations between single protein and different diseases differed. 111 showed directionally concordant associations with all disease types, including positive associations for 83 proteins and inverse associations for 28 proteins. In contrast, 22 proteins showed opposite associations with different diseases (i.e. positive associations with some and inverse associations with the others).

Genetic associations between proteins and CMD

The evaluation of MR evidence included 35 studies assessing circulating proteins as possible causal biomarkers for CMDs, with 10,531 protein-disease pairs reported and 1614 pairs eligible for meta-analysis. Different from the observational studies, the genetic associations between proteins and six stroke subtypes were investigated, including total stroke, IS, large artery stroke (LAS), cardioembolic stroke (CES), small vessel stroke (SVS), haemorrhagic stroke (HS), and subarachnoid haemorrhage (SAH). The certainty of evidence derived from MR studies were divided into four tiers, and 245 proteins were graded as Tier 1 and Tier 2 (Figure 1-2 and Figure S2). Moderate heterogeneity was observed for MR pooled results, and 14.2% pairs had $I^2 \geq 80\%$. Detailed effect estimates for each disease were summarized in Table S9-S21.

When comparing the observational and genetic associations in the same study, 39 of 246 protein-disease pairs (15.8%) showed consistent results (Table S22). Of 1731 protein-disease pairs

investigated in both of observational and MR pooled analyses, only 22 pairs showed directionally consistent associations (i.e. satisfying significant observational associations and Tier 1-2 proteins on MR evidence, Figure 2).

Of the 35 proteins significant in the meta-analysis of observational studies for CHD, only MMP12 was Tier 1 or 2 targets in MR studies (Figure 2), but the directions of associations were inconsistent with observational studies (OR, 1.29; 95% CI, 1.09-1.52; *p*-value=0.003) and MR studies (OR, 0.97; 95% CI, 0.94-1.00; *p*-value=0.022).

Within the set of 31 proteins exhibiting significance in observational results for IS, ADML and MMP12 were also identified as Tier 1 or 2 targets (Figure 2). ADML and MMP12 were associated with higher risk of IS in observational meta-analysis, while both of them were associated with lower risk of IS in MR studies (Figure 3).

Among the 323 proteins found significant in observational studies for T2D, 15 proteins belonged to Tier 1 or 2 targets (Figure 2). Nine proteins showed directionally consistent associations with risk of T2D between observational and MR studies, and the remaining six proteins showed opposite associations (Figure 3).

In the set of 286 HF-associated proteins identified in the meta-analysis of observational studies, the MR evidence of 27 proteins were graded as Tier 1 or 2 (Figure 2). The results of 12 proteins were directionally consistent in observational and MR analyses, and the results of 15 proteins were directionally opposite (Figure 3).

There were 57 proteins significant in the observational results for AF, among which only three proteins were classified with Tier 1 or 2 MR evidence (Figure 2). SPON1 was directionally

consistent (RR, 1.37; 95% CI, 1.11-1.69 in observational studies vs. OR, 1.08; 95% CI, 1.02-1.15 in MR studies); the remaining two were directionally inconsistent, namely FBLN3 (RR, 1.80; 95% CI, 1.50-2.17 in observational studies vs. OR, 0.94; 95% CI, 0.90-0.97 in MR studies) and LEP (RR, 0.90; 95% CI, 0.81-1.00 in observational studies vs. OR, 1.14; 95% CI, 1.00-1.29 in MR studies).

Of the 16 proteins significantly associated with atherosclerosis in observational studies, only ANFB was considered as Tier 1 or 2 in MR pooled results (Figure 2), which was inversely associated with risk of atherosclerosis in MR studies (β, -0.006; 95% CI, -0.009--0.003; p-value=4.40×10^{-5}), but showed positive association in observational pooled results (β, 0.006; 95% CI, 0.001-0.010; p-value=0.014).

Combining the associations between a single protein and various CMDs, we identified 23 Tier 1-2 proteins associated with risk of two or more CMDs, referred to as "pleiotropic protein" (Figure S2). These included 14 proteins associated with two diseases, three proteins with three diseases, three proteins (TMPS5, TNF12, TNR5) with four diseases, one protein (LPA) with six diseases, and two proteins (IL6RA and MMP12) associated with seven diseases. The directions and strengths of associations between single protein with different diseases differed. Of these 23 proteins, 18 showed directionally concordant associations with all disease types, including positive associations for eight proteins (LPA, BGAT, FGF5, HSPB1, I15RA, MMP3, NELL1, TMPS5) and inverse associations for ten proteins (CATD, DHPR, ERAP1, FCG2A, IL6RA, MMP12, QSOX2, SCAR5, TFPI1, TNR5). In contrast, five proteins showed directionally opposite associations with different diseases (CFAI, IL1R2, MANBA, SPON1, TNF12).

Quality assessment

For observational studies, 54 studies (90.0%) were graded as high quality, and the remaining six studies with an NOS score ≤six were considered as moderate quality (Table S23). Table S24 summarizes the validation of three assumptions by each MR study. Assumption one was validated in eight studies (22.3%), and assumptions two and three were verified in 12 studies (34.3%). All three assumptions were validated in six studies (17.7%). To reduce bias due to pleiotropy, 29 studies (82.9%) restricted instrumental variables to *cis*-pQTLs, and 19 studies (54.3%) employed MR-Egger regression as sensitivity analysis, of which 14 studies (73.7%) reported no significant signs of horizontal pleiotropy (Table S24).

Risk prediction models including proteins

29 studies constructed risk prediction models for incident CMD including proteins and compared that with clinical risk models (Table S25). The number of proteins included in the model ranged from 1 to 291 (median 6, interquartile range 1-20). Although protein models showed better discrimination over the clinical risk model, the majority had limited improvement (Figure 5). 13 out of 79 models (16.5%) improved the C-index by ≥0.10 and 32 models (40.5%) reported significant improvement. There were 28 models (35.4%) with proteins reaching a C-index ≥0.8, half of which had a base model without proteins with a C-index <0.8. The most commonly included proteins were ANFB (19 models) and IBP2 (9 models). Disease outcomes with C-index improvement ≥0.10 were arteriosclerotic cardiovascular disease (ASCVD) (n=2), T2D (n=1), and HF (n=10). In the top two models that improved the C-index for predicting HF (difference of C-index≈0.19), both included PRELP, LEG9, NEMO, and UPAR.

Evaluation of druggability and clinical development activity

Of the proteins identified as Tier 1 or 2, 102 (41.6%) were established drug targets in the database (Table S26). These included three target-indication pairs (CP3A4-HF, LYAM2-HF, and PLMN-HF) that had already been approved as treatments. There were no reports of drug targets or drug development for the remaining 288 protein-disease pairs. Additionally, a total of 66 Tier 1 or 2 proteins were targets of licensed drugs for indications different from the diseases implicated by our MR pooled results (Figure S3).

Functional annotation and enrichment analysis

Of the proteins identified as Tier 1 or 2, four, eight, 19, 85, and 12 GO biological processes identified for CHD, IS, T2D, HF, and AF (p-value<0.05), and nine processes were related to ≥two diseases (Table S27). The top 20 GO biological processes (i.e. terms with lowest p-value) were shown in Figure 6A. There were five, two, six, 11, and two KEGG pathways identified for CHD, IS, T2D, HF, and AF (p-value <0.05), respectively, and four pathways were related to ≥two diseases (Table S28, Figure 6B).

Discussion

In the current study, 560 proteins were observationally associated with CMD (including 133 proteins associated with ≥two CMD subtypes), while 245 proteins showed genetic associations with CMD (including 23 proteins showing pleiotropic effects). 22 protein-disease pairs showed directionally consistent associations in observational and MR pooled estimates. 288 Tier 1 or Tier 2 protein-disease pairs were not reported for drug development and 66 proteins were drug targets approved for other indications, providing new possible targets for drug development and repurposing opportunities for existing drug targets. Addition of proteins to a clinical factor model

modestly improved risk prediction for incident CMD.

Several proteins showed consistent associations with diseases in both observational and MR analyses (e.g. TNF12), whereas some yielded inconsistent results (e.g. MMP12). TNF12 (also known as TNFSF12) was inversely associated with CES and AF in MR studies, and with HF in observational studies. TNF12 has been investigated in phase I-II trials for lupus nephritis, rheumatoid arthritis and neoplasms, but not for CMD.[8] MMP12 plays an important role in maintaining vein wall structure and function.[9] The pooled MR results revealed that MMP12 was associated with lower risk of CHD, IS, LAS, CES, T2D, and HF, but a higher risk of HS. In contrast, observational meta-analysis found MMP12 positively associated with risk of CHD, HF, and IMT-CCA. As a therapeutic target, lithostat, a MMP12 inhibitor, is used to treat urea splitting bacterial infections of the urinary tract[8], and two other MMP12 inhibitors (i.e. neovastat, marimastat) did not improve cancer survival in phase III trials.[10,11] No CMD-related drug development for MMP12 was found. The heterogeneity between observational and MR studies might be partly explained by confounding and reverse causality in observational studies[12]and the validity of MR assumptions[13].

Our findings suggested that a targeted proteomics panel might improve CMD risk prediction. 32 of 79 models included in this study showed better performance of the protein model over the conventional clinical model. However, only 13 of these models improved the C-index by ≥0.10. Previous studies also indicated that protein risk scores[14] and polygenic risk scores[15] (also applying omics data to CMD risk prediction models) both provided statistically significant but modest improvement in discrimination. The protein risk score increased the C-index by 0.014 for ASCVD

prediction[14], while the polygenic risk score improved the C-index by 0.024 for CHD prediction[15]. Due to the relatively high economic cost of high-throughput proteomics tests and the heterogeneity of proteins proposed by different studies, it remained to be cautiously determined whether protein biomarkers are clinically useful to screen for future CMD.[16]

The current study identified 288 protein-disease pairs to be putative causal biomarkers but without evidence of drug development, representing potential new therapeutic targets for CMD. This study also observed putative drug-repurposing opportunities of some existing drugs. For example, PAR1 protein is targeted by vorapaxar[17], which is used to reduce thrombotic cardiovascular events in patients with history of myocardial infarction or peripheral arterial disease.[18] Our study showed both observational and genetic evidence supporting the role of inhibition of circulating level of PAR1 on reducing HF risk, implying a repurposing opportunity of vorapaxar on HF prevention.

This study also had several limitations. First, most included studies used high-throughput, targeted proteomics platforms covering 85-5000 proteins, with varying protein content. These platforms selected proteins related to cardiometabolic health and other factors based on hypotheses drawn from previous research. However, it's possible that some unmeasured proteins may still be associated with CMDs. Second, we did not perform subgroup analyses by regions or race/ethnicity as the number of studies in non-European populations was limited. Third, participants included in the meta-analysis were mostly Europeans, and the results might not be directly generalizable to populations with different ethnic/racial backgrounds. Fourth, although our results indicated potential causal roles of certain proteins in CMD, the restriction of biological samples to blood specimens indicated that the results did not specifically address in which tissue the effects may be

mediated. Fifth, horizontal pleiotropy (i.e. instrumental variables additionally influence the outcome through pathways that bypass the exposure[19]) is a major consideration and limitation to MR studies. We conducted a quality assessment of MR studies, focusing on addressing horizontal pleiotropy, and only included *cis*-loci in the meta-analysis to minimize potential bias. Lastly, although adherence to Hardy-Weinberg equilibrium was used to control potential genotyping errors in each MR study included, establishing causality is challenging due to the heterogeneity in study designs and proteomics coverage. Therefore, we employed several approaches to identify significant proteins and enhance results credibility, including meta-analysis, MR evidence grading, and druggability evaluation.

In conclusion, this study comprehensively integrated evidence on the observational and genetic associations of proteins with CMD, revealing the important roles of circulating proteins in CMD. The identification of novel protein biomarkers offered promising targets for drug development and risk stratification. These findings enhanced our understanding of CMD aetiology and highlighted the potential of circulating proteins as biomarkers and therapeutic targets, paving the way for future research and clinical applications.

Contributors

Y.P., T.W., and Y.K. conceived and designed research; T.W. and Y.K. collected, analysed and interpreted data; T.W. and Y.K. drafted the manuscript; Y.P. revised the paper; Y.L., Z.W., J.L., C.Y., D.S., P.Y., C.K., Z.C., and L.L. contributed to significant amendments to the final manuscript; Y.P. had primary responsibility for final content. All authors read and approved the final manuscript.

Ethical approval and consent to participate

Not applicable.

Data sharing statement

The data underlying this article are available in the article and in its online supplementary material.

Declaration of interests

We declare no competing interests.

Funding

This work was supported by National Natural Science Foundation of China (82304223, 82192904, 82192901, 82192900). The funders had no role in the study design, data collection, data analysis and interpretation, writing of the report, or the decision to submit the article for publication.

Acknowledgements

None

References

1. GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 2020;**396**:1223-1249. doi: 10.1016/s0140-6736(20)30752-2
2. Tahir UA, Gerszten RE. Omics and Cardiometabolic Disease Risk Prediction. *Annu Rev Med* 2020;**71**:163-175. doi: 10.1146/annurev-med-042418-010924
3. Suhre K, McCarthy MI, Schwenk JM. Genetics meets proteomics: perspectives for large population-based studies. *Nat Rev Genet* 2021;**22**:19-37. doi: 10.1038/s41576-020-0268-2
4. Sanderson E, Glymour MM, Holmes MV, Kang H, Morrison J, Munafò MR*, et al.* Mendelian randomization. *Nature Reviews Methods Primers* 2022;**2**:6. doi: 10.1038/s43586-021-00092-5
5. Holmes MV, Richardson TG, Ference BA, Davies NM, Davey Smith G. Integrating genomics with biomarkers and therapeutic targets to invigorate cardiovascular drug development. *Nature Reviews Cardiology* 2021;**18**:435-453. doi: 10.1038/s41569-020-00493-1
6. The UniProt, Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* 2023;**51**:D523-D531. doi: 10.1093/nar/gkac1052
7. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;**4**:44-57. doi: 10.1038/nprot.2008.211
8. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR*, et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;**46**:D1074-d1082. doi: 10.1093/nar/gkx1037
9. Olejarz W, Łacheta D, Kubiak-Tomaszewska G. Matrix Metalloproteinases as Biomarkers of Atherosclerotic Plaque Instability. *Int J Mol Sci* 2020;**21**. doi: 10.3390/ijms21113946
10. Escudier B, Choueiri TK, Oudard S, Szczylik C, Négrier S, Ravaud A*, et al.* Prognostic factors of metastatic renal cell carcinoma after failure of immunotherapy: new paradigm from a large phase III trial with shark cartilage extract AE 941. *J Urol* 2007;**178**:1901-1905. doi: 10.1016/j.juro.2007.07.035
11. Shepherd FA, Giaccone G, Seymour L, Debruyne C, Bezjak A, Hirsh V*, et al.* Prospective, randomized, double-blind, placebo-controlled trial of marimastat after response to first-line chemotherapy in patients with small-cell lung cancer: a trial of the National Cancer Institute of Canada-Clinical Trials Group and the European Organization for Research and Treatment of Cancer. *J Clin Oncol* 2002;**20**:4434-4439. doi: 10.1200/jco.2002.02.108
12. Sekula P, Del Greco MF, Pattaro C, Köttgen A. Mendelian Randomization as an Approach to Assess Causality Using Observational Data. *J Am Soc Nephrol* 2016;**27**:3253-3265. doi: 10.1681/asn.2016010098
13. Hu X, Zhao J, Lin Z, Wang Y, Peng H, Zhao H*, et al.* Mendelian randomization for causal inference accounting for pleiotropy and sample structure using genome-wide summary statistics. *Proceedings of the National Academy of Sciences* 2022;**119**. doi: 10.1073/pnas.2106858119
14. Helgason H, Eiriksdottir T, Ulfarsson MO, Choudhary A, Lund SH, Ivarsdottir EV*, et al.* Evaluation of Large-Scale Proteomics for Prediction of Cardiovascular Events. *Jama* 2023;**330**:725-735. doi: 10.1001/jama.2023.13258
15. Patel AP, Wang M, Ruan Y, Koyama S, Clarke SL, Yang X*, et al.* A multi-ancestry polygenic risk score improves risk prediction for coronary artery disease. *Nat Med* 2023;**29**:1793-1803. doi: 10.1038/s41591-023-02429-x
16. Nurmohamed NS, Kraaijenhof JM, Mayr M, Nicholls SJ, Koenig W, Catapano AL*, et al.* Proteomics

and lipidomics in atherosclerotic cardiovascular disease risk prediction. *Eur Heart J* 2023;**44**:1594-1607. doi: 10.1093/eurheartj/ehad161

17. Bonaca MP, Morrow DA. SCH 530348: a novel oral thrombin receptor antagonist. *Future Cardiol* 2009;**5**:435-442. doi: 10.2217/fca.09.27
18. Tricoci P, Huang Z, Held C, Moliterno DJ, Armstrong PW, Van de Werf F*, et al.* Thrombin-receptor antagonist vorapaxar in acute coronary syndromes. *N Engl J Med* 2012;**366**:20-33. doi: 10.1056/NEJMoa1109719
19. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014;**23**:R89-98. doi: 10.1093/hmg/ddu328

Figure legends

Figure 1. Overview of analytic approaches and key findings. *In four observational studies, incident ischemic stroke was examined as the outcome and no other stroke subtypes were investigated. AF, atrial fibrillation; CHD, coronary heart disease; CMD, cardiometabolic disease; HF, heart failure; MR, Mendelian randomization study; pQTL, protein quantitative trait loci; T2D, type 2 diabetes.

Figure 2. Summary of proteins in observational and MR studies. The number of proteins in observational and MR pooled results following three steps (i.e. reported by original studies, included in meta-analysis, and significantly associated with CMD in observational meta-analysis or with Tier 1 or 2 evidence in MR pooled results), and comparison of proteins significant in observational meta-analysis and Tier 1 or 2 proteins in MR pooled results. Consistent proteins denote protein biomarkers showing observational and genetic associations in the same direction, while inconsistent proteins denote protein biomarkers showing opposite associations in observational and genetic analyses. AF, atrial fibrillation; CHD, coronary heart disease; HF, heart failure; IS, ischemic stroke; MR, Mendelian randomization study; OB, observational study; T2D, type 2 diabetes.

Figure 3. Associations between proteins and CMDs in observational meta-analysis and MR pooled results. Names were given for top 20 proteins with the lowest *p*-value. AF, atrial fibrillation; CHD, coronary heart disease; HF, heart failure; IS, ischemic stroke; T2D, type 2 diabetes.

Figure 4. Pleiotropy of proteins in observational pooled results. Grey colour denotes that the protein-disease pair was not available for meta-analysis. AF, atrial fibrillation; CHD, coronary heart disease; HF, heart failure; IS, ischemic stroke; T2D, type 2 diabetes.

Figure 5. Improvement of C-index of clinical risk prediction model through the addition of proteins. The size of the dots represents the number of proteins included in the model. The colour of the dots indicates the C-index of the clinical model and the model with addition of proteins: blue denotes that both the base model and the protein model have a C-index <0.8; yellow denotes that the base model has a C-index <0.8, and the protein model has a C-index ≥0.8; red denotes that both the base model and the protein model have a C-index ≥0.8. An improvement in C-index ≥0.1 is considered clinically meaningful. AF, atrial fibrillation; ASCVD, atherosclerotic cardiovascular disease; HF, heart failure; T2D, type 2 diabetes.

Figure 6. Chord diagrams of enriched in GO biological processes and KEGG pathways for CMD. A shows the top 20 GO biological processes and B shows significant KEGG pathways. AF, atrial fibrillation; CHD, coronary heart disease; HF, heart failure; IS, ischemic stroke; T2D, type 2 diabetes.