Plasma Proteomics Linking Primary and Secondary diseases: Insights into Molecular Mediation from UK Biobank Data

Hanyu Qian[a,b,*], Chao Wu[a,*], Bo Li[a], Anthony Rosenzweig[a,#], Meng Wang[a,b,#]

a. Stanley and Judith Frankel Institute for Heart and Brain Health, University of Michigan Medical Center, Ann Arbor, MI, USA
b. Gilbert S. Omenn Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA
*. Joint first author
#. Corresponding authors: anthros@med.umich.edu; mwmeng@umich.edu

Abstract

Diabetes, hypertension, and dyslipidemia are major risk factors for cardiovascular, neurological, renal, and pulmonary diseases. However, the molecular mediators linking these primary cardiometabolic diseases to downstream disease development, particularly clinically accessible biomarkers, remain incompletely understood. Here, we investigated the mediating roles of plasma proteins in secondary disease development using data from 53,030 UK Biobank participants. Across three primary diseases and 18 subsequent conditions, we identified 1,461 significant mediation pathways involving 395 unique plasma proteins. Notable examples include GDF15 consistently mediating the diabetes–cardiovascular disease link and ADM mediating the hypertension–pulmonary disease pathway. Identified mediating proteins are highly enriched in receptor-mediated signaling and molecular interaction pathways. Mendelian randomization supports causal roles for 84 proteins, highlighting their potential as therapeutic targets. Moreover, these protein mediators improved predictive accuracy for secondary disease risk compared with models using other plasma proteins or traditional clinical risk factors from machine learning methods. For example, among individuals with hypertension, the inclusion of top mediating proteins improved prediction accuracy for glomerular disease risk by approximately 14% measured by C-index. Stratification by biomarker-defined levels of primary disease control and severity further reveals additional disease pathways and protein mediators, including APOE linking poorly controlled diabetes to increased risk of Alzheimer's disease. These findings implicate plasma proteins as key molecular mediators connecting cardiometabolic diseases to

their downstream complications and nominate promising targets for biomarker development and therapeutic intervention.

Introduction

With global populations expanding and aging, alongside the rising epidemic of metabolic disorders, the burden of chronic diseases is escalating at an unprecedented rate. Diabetes, hypertension, and dyslipidemia are among the most prevalent and consequential chronic health conditions worldwide. According to the World Health Organization (WHO), an estimated 1.28 billion adults are affected by hypertension, 830 million are living with diabetes, and approximately 39% of the global adult population exhibit elevated blood cholesterol levels[1]. These cardiometabolic chronic conditions, hereafter referred to as primary diseases, not only exert substantial economic and social burdens but also serve as major drivers of downstream health complications[2,3]. For example, diabetes and hypertension are well-recognized drivers of multi-organ disease development, contributing to elevated risks of cardiovascular outcomes—including heart failure (HF) and ischemic heart disease (IHD) [4–7] —as well as renal dysfunction and chronic kidney disease [8].

Previous evidence has identified a range of biological mediators linking metabolic and hemodynamic disorders (i.e. diabetes mellitus, hypertension, and dyslipidemia), to downstream pathological processes across multiple organ systems. These mediators encompass chronic inflammation, oxidative stress, endothelial dysfunction, altered adipokine signaling, and broader metabolic dysregulation. Pro-inflammatory signals, such as interleukin-6 (IL-6) and monocyte chemoattractant protein-1 (MCP-1) are consistently upregulated in metabolic disease and contribute to endothelial activation, monocyte recruitment, and extracellular matrix remodeling[9]. Concurrently, renin-angiotensin-aldosterone system (RAAS) activation enhances oxidative stress via NADPH oxidase, activates NF-κB in vascular and immune cells, leading to upregulation of inflammatory cytokines, chemokines, and adhesion molecules. These pathways accelerate hypertension, vascular inflammation, and atherosclerosis, especially in the context of diabetic and dyslipidemia[10,11]. Understanding these mediators not only advances mechanistic insight into the metabolic disease continuum but also highlights novel therapeutic targets beyond traditional risk factor management. However, many mediators identified to date may be challenging to assess routinely in clinical practice due to limitations in measurement techniques or biomarker

availability. Moreover, the broader landscape of clinically accessible mediators detectable in blood remains largely unexplored. A systematic characterization of such biomarkers could substantially improve risk stratification, refine preventive strategies, and facilitate the development of targeted interventions, particularly for populations facing already high burdens of disease.

Emerging high-throughput plasma proteomic platforms, including Olink[12] and SomaScan[13], enable the quantification of thousands of plasma proteins from minimally invasive blood samples, providing unique insights into systemic biology. Plasma proteomics has shown promise in estimating organ biological aging[14,15], predicting incident disease and mortality[16–18], and linking diverse phenotypic traits to underlying disease status[19,20]. Leveraging these advances, we applied large-scale plasma proteomic profiling within a rigorous mediation analysis framework[21] to systematically investigate the mediating roles of circulating proteins during disease development.

In this study, we used data from 53,030 UK Biobank participants (aged 40–69 years at baseline) with plasma proteomic measurements and longitudinal follow-up of clinical outcomes[22,23] to identify mediating proteins in disease pathways linking three prevalent conditions—diabetes, hypertension, and dyslipidemia—to 18 secondary disease outcomes across four major clinical domains (Fig. 1). We further annotated their biological functions, assessed their potential causal effects on secondary diseases using Mendelian randomization, and evaluated their value in predicting secondary disease risk among individuals with primary diseases. We also stratified participants by levels of primary disease control and severity, defined using continuous clinical biomarkers, to explore additional disease pathways and protein mediators.

Results

Population Characteristics and Diseases

A total of 53,030 participants were in the study, with an average age of 57.4 years, of whom 53.9% were women (Supplementary Table 1). After data preprocessing (Methods), expression data for 2,922 plasma proteins and clinical data on three primary diseases--diabetes, hypertension, and dyslipidemia (Supplementary Table 2), as well as 18 secondary diseases (Table 1 and Supplementary Table 3), were used for analysis to investigate candidate proteins

involved in the primary-secondary disease pathway[23]. In this study, primary diseases were defined as chronic cardiometabolic conditions that increase the risk of subsequent health outcomes, while secondary diseases were defined as potential downstream outcomes associated with the presence of primary diseases. Primary disease status was determined based on whether participants were diagnosed with a primary disease within the period spanning two years before to one year after the plasma protein data collection. The number of positive cases per primary disease ranged from 1,053 (2.0%) to 3,012 (5.7%).

We examined 18 secondary diseases spanning four major organ systems: seven cardiovascular diseases (CVD), two renal diseases (RD), two pulmonary diseases (PuDs), and seven neurological diseases (NDs) (Table 1 and Supplementary Table 3). These domains were selected based on extensive evidence that the three primary cardiometabolic conditions—diabetes, hypertension, and dyslipidemia—exert systemic effects on the heart, kidney, lung, and brain[3,24–28]. Secondary disease outcomes were further selected according to the UK Biobank disease classification framework and required to have sufficient prevalence to ensure statistical power. Incident case counts, defined as diagnoses occurring after plasma protein collection, ranged from 106 (0.2%) to 8,962 (16.9%) across secondary diseases (Supplementary Table 3).

Identification of Diabetes-Associated Plasma Protein Mediating Pathways

Diabetes was identified as a significant risk factor for six out of 18 secondary diseases, as indicated by hazard ratios (HRs) greater than 1 in Cox regression models (Fig. 2a, Methods). The strongest associations were observed in glomerular diseases (GD; HR = 2.501, $p = 9 \times 10^{-4}$), and renal tubulo-interstitial disease (RTID; HR = 2.034, $p < 1 \times 10^{-10}$), with both GD and RTID classified as renal diseases, consistent with prior reports[29–31]. Diabetes was also significantly associated with the risk of four CVDs, including cerebrovascular diseases (CeD), cardiac arrest and arrhythmias (CAA), artery diseases (ArD), and HF, supporting previous studies[3,32].

Among the 2,922 plasma proteins analyzed, 1260 were significantly associated with diabetes (with Benjamini-Hochberg [BH] adjusted p values[33] < 0.05, Fig. 2b). Several strong associations emerged including the association with GDF15, ADGRG1, and NFASC ($p < 1 \times 10^{-10}$ for all). Notably, GDF15, a well-established metabolism-related protein, has been repeatedly linked to diabetes in prior studies[34], supporting the robustness of our findings.

Significant diabetes-associated plasma proteins and secondary diseases were further examined under a mediation analysis framework[21] to assess whether circulating proteins at least partially mediate the development from diabetes to its secondary outcomes (Fig. 1c, Methods). To ensure robustness, we applied three criteria to identify mediators and their associated mediating disease pathways: (i) BH adjusted $p < 0.05$ for the indirect effect; (ii) An absolute indirect effect ≥ 0.01 (in terms of the log hazard ratio); and (iii) A proportional indirect effect (PIE), defined as the ratio of the indirect effect to the total effect, ≥ 10%. In total, 1,057 significant diabetes–associated plasma protein mediating pathways were identified, involving 341 unique plasma proteins among six secondary diseases (Fig. 2d and Supplementary Table 4). As expected, all indirect effects were positive, with HRs greater than 1, indicating that diabetes potentially increased the risk of secondary diseases via protein-mediated mechanisms. Notably, 25 proteins, including GDF15, LGALS4, PRSS8, and HAVCR1, consistently mediated pathways linking diabetes to all six secondary diseases (Fig. 2d; Supplementary Fig. 2a). Many of these high-frequency mediators also exhibited strong and recurrent indirect effects across multiple disease pathways (Fig. 2e; Supplementary Table 5). Among them, GDF15, a well-established biomarker previously linked to diabetes[35,36], emerged as the most prominent mediator, consistently showing the largest indirect effects across all six diabetes-related mediation pathways. In addition, LGALS4 consistently ranked among the top five mediators by indirect effect size across diabetes-associated mediation pathways, consistent with previous observational studies reporting associations between circulating LGALS4 levels and diabetes-related traits[37]. The ranking of the top five mediator proteins remained largely consistent in sensitivity analyses incorporating 14 additional covariates into the mediation framework, including the Townsend deprivation index (TDI), alanine aminotransferase, urea, triglycerides, and related clinical factors, with an average Jaccard index[38] 0.64 compared with the primary model across identified disease pathways (Methods, Supplementary Fig. 5).

Because glycemic control is a key determinant of diabetes-related complications, we further examined whether glycated hemoglobin (HbA1c) levels were associated with differential risks of secondary diseases and distinct protein-mediated pathways. Participants were stratified according to HbA1c levels[39]: normal (HbA1c < 5.7%), prediabetes (5.7% ≤ HbA1c < 6.4%), and severe diabetes (HbA1c ≥ 6.4%). This stratification revealed six additional disease pathways not identified when using diagnosis history alone (Fig. 2f), including pathways from diabetes to

three CVDs, IHD, myocardial disorder (MyD), and valve disorder (VD); two NDs (Alzheimer's disease [AD], and nerve root and plexus disorders [NRPD]); and one PuD (other respiratory diseases principally affecting the interstitium, [OPDRI]). These findings suggest that poor glycemic control, even during the prediabetes stage, may increase the risk of downstream diseases. Participants with severe diabetes also exhibited significantly higher risks of AD, consistent with clinical observations[40].

Although most key protein mediators identified using HbA1c stratification overlapped with those identified using diagnosed history for the same secondary disease (Supplementary Fig. 4c, Supplementary Table 10 and Supplementary Table 11), notable differences were observed (Fig. 2g). For example, GDF15 and LGALS4 exhibited significant indirect effects across 21 disease pathways, the largest number of mediated pathways among all plasma proteins. However, GDF15 showed strong indirect effects primarily in severe diabetes and diabetes diagnosed groups, but not in the prediabetes group (Fig. 2h). In contrast, LGALS4 consistently acted as a strong mediator across all levels of glycemic control. Additional protein mediators emerged in pathways specific to stratified HbA1c control levels. For instance, APOE, PVR, and NEFL were identified in the pathway leading to AD, with APOE being a well-established biomarker associated with neurological disorders[41]. Similarly, CFI, CFH, and SERPIND1 emerged as strong mediators in the pathways leading to IHD, highlighting additional molecular links between glycemic dysregulation and cardiovascular risk. We also evaluated optimal HbA1c control levels associated with increased disease risk. For AD, we identified two critical HbA1c "acceleration points" at approximately 5.7% and 7.9%, beyond which AD risk increases substantially, underscoring the importance of maintaining HbA1c within the normal range to mitigate downstream secondary disease risk (Supplementary Materials).

Identification of Hypertension-Associated Plasma Protein Mediating Pathways

We applied the mediation analysis framework to investigate the hypertension-related secondary diseases and their potential protein mediators (Methods). Hypertension was identified as a significant risk factor for eight out of 18 secondary diseases in this study (Fig. 3a). The strongest association was observed between hypertension and GD (HR = 2.336, $p = 0.002$). Hypertension was significantly associated with 928 plasma proteins (BH adjusted p values < 0.05, Fig. 3b), with the strongest association observed for HAVCR1 ($p < 1 \times 10^{-10}$).

Among these 928 hypertension-associated plasma proteins, 64 unique proteins exhibited significant indirect effect in at least one disease pathway (Fig. 3c). GDF15 and HAVCR1 emerged as key mediators, mediating six and five secondary disease associations, respectively (Fig. 3d, Supplementary Fig. 2c). The HAVCR1 gene encodes the immunoglobulin-mucin protein TIM-1/KIM-1, which is shed from injured proximal tubular cells and is currently recognized as an FDA-qualified, ultra-early biomarker for predicting both acute kidney injury and long-term decline in estimated glomerular filtration rate (eGFR)[42]. Another notable example was ADM which ranked among top five proteins mediating the associations between hypertension and four secondary diseases (Fig. 3e). ADM is markedly upregulated as a compensatory vasodilatory and natriuretic peptide that counterbalances excessive vascular tone, attenuates fibrosis, and preserves renal and cardiac function[43]. Elevated ADM levels have also been strongly associated with adverse cardiovascular and renal outcomes[44], suggesting that the observed indirect effects may reflect sustained blood pressure elevation and compensatory vascular responses rather than direct promotion of disease development. Sensitivity analyses adjusting for additional covariates yield consistent results, with an average Jaccard index 0.66 compared with the primary model across identified disease pathways (Methods).

We further stratified participants according to systolic and diastolic blood pressure (SBP and DBP) into four categories: normal (< 140/90 mmHg), hypertension grade 1 (140–159/90–99 mmHg), hypertension grade 2 (160–179/100–109 mmHg), and hypertension grade 3 (≥ 180/110 mmHg), according to the 2023 European Society of Hypertension (ESH) guidelines[45] (Methods). Using this stratification, we identified eight significant disease pathways. In each pathway, participants in at least one hypertension grade, consistently including grade 3, exhibited a significantly higher risks of developing secondary diseases compared to those with normal blood pressure (Fig. 3f). Seven of these eight secondary diseases overlapped with those identified using hypertension diagnosis history. The only additional disease development pathway was ArD among participants with grade 3 hypertension (HR = 1.401, 95% CI: 1.052–1.866), which was not observed among participants with grades 1 or 2 hypertension nor in analyses using diagnosis history. As expected, secondary diseases associated with milder hypertension were consistently retained among pathways identified at higher hypertension grades, with risk magnitude generally increasing with hypertension severity. Overall, the risk profile of participants with diagnosed hypertension generally resembled those with grade 2-3 hypertension.

From the mediation analysis, in total, 197 plasma proteins exhibited significant indirect effects across pathways from at least one hypertension grade group, with grade 3 hypertension displaying the broadest and strongest mediation signals (Supplementary Fig. 6b). Several mediators overlapped with those identified using diagnosis history, including HAVCR1, MMP7, and NT-proBNP (Fig. 3f and 3g), in which NT-proBNP is a well-established biomarker of myocardial stress and heart failure[46].

Identification of Dyslipidemia-Associated Plasma Protein Mediating Pathways

Dyslipidemia showed the broadest downstream clinical impact, with significant associations with ten out of 18 secondary diseases (Fig. 4a). In total, 557 circulating plasma proteins were significantly associated with dyslipidemia (Fig. 4b). Among these, PCSK9 showed the strongest association ($p < 1 \times 10^{-10}$), consistent with its well-established role in lipid metabolism and clinical intervention[47]. PCSK9 inhibitors, including monoclonal antibodies (alirocumab, evolocumab) and siRNA-based therapy (inclisiran), are approved adjuncts to statins and have demonstrated substantial efficacy in lowering LDL cholesterol levels and cardiovascular event rates in patients with familial hypercholesterolemia or established atherosclerotic cardiovascular disease (ASCVD)[47].

Under the same mediation analysis framework, we identified 131 unique plasma proteins with significant indirect effects linking dyslipidemia and ten secondary diseases (Fig. 4c). LGALS4 and PRSS8 exhibited significant indirect effects in eight and seven pathways, respectively, representing the highest numbers among all dyslipidemia-associated mediators (Fig. 4d, Supplementary Fig. 2b). Notably, LGALS4 consistently ranked among the top five mediators by indirect effect across multiple dyslipidemia-associated pathways, suggesting a broader functional role in metabolic disease development. In addition to these shared mediators, FABP1, a key protein involved in fatty acid metabolism[48], emerged as a prominent dyslipidemia-specific mediator (Fig. 4e), underscoring its potential relevance to lipid-driven disease pathway. FABP1 acts at the interface of hepatic lipid handling and systemic metabolic homeostasis[29]. By regulating fatty acid trafficking and cholesterol metabolism, FABP1 influences the development of dyslipidemia, which in turn accelerates atherosclerosis, vascular dysfunction, and cardiac remodeling[49]. Elevated circulating FABP1 has been consistently linked to obesity, insulin

resistance, and cardiovascular events[50], raising the possibility that it operates as both a mechanistic mediator of lipid-driven pathology and a sensitive biomarker of hepatic stress.

We further identified 31 plasma proteins that exhibited significant indirect effects in pathways linking all three primary diseases to CAA, RTID, or HF, three secondary diseases whose risks were significantly influenced by all primary conditions (Fig. 4f and 4g; Supplementary Fig. 3). Notably, GDF15 and IGFBP4 mediated associations across all three primary diseases and each of these secondary diseases, highlighting their central role in shared multimorbidity-related pathways. In addition, ACE2, a well-established therapeutic target in renal disease[51], specifically mediated associations between all three primary diseases and RTID.

Biological Function Annotation of the Protein Mediators

To understand the biological functions underlying cross-disease mediation, we performed enrichment analysis using the Gene Ontology Biological Process (GO:BP)[52,53] databases to annotate enriched biological processes and pathways[54] (Methods). Across all disease pathways, significant mediators showed consistent enrichment in pathways related to receptor activity and molecular binding functions (Fig. 5a and Supplementary Table. 12). Mediators linking diabetes with arterial disease, cardiac arrest and arrhythmias (CAA), and glomerular diseases were enriched for signaling receptor activity, together with pathways involving carbohydrate, organic acid, and glycosaminoglycan binding, indicating shared receptor-mediated molecular interactions across vascular, cardiac, and renal complications. Similarly, mediators linking diabetes with heart failure and renal tubulointerstitial diseases showed enrichment for death receptor activity, growth factor binding, receptor ligand activity, and cytokine binding, suggesting involvement of apoptotic and ligand–receptor signaling pathways. In contrast, mediators linking dyslipidemia and hypertension with renal diseases were primarily enriched for fatty acid and organic acid binding pathways, while only a single pathway reached significance for hypertension-heart failure mediation. These findings indicate that receptor-mediated signaling and molecular interaction pathways represent dominant biological processes underlying protein-mediated disease-to-disease relationships.

Mendelian Randomization Reveals Potential Causal Relationships Between Protein Mediators and Secondary Diseases

We applied a Mendelian randomization[55] (MR) framework[56,57] to investigate whether plasma proteins exhibiting significant indirect effects also play causal roles in the development of secondary diseases. In this framework, protein expression was treated as the exposure, and secondary disease incidence as the outcome, with genome-wide significant *cis*-protein quantitative trait loci (*cis*-pQTLs) as instrumental variables[23]. We employed a two-sample MR framework utilizing 12 methods to account for pleiotropy and heterogeneity, supplemented by Steiger directionality testing to confirm exposure-to-outcome effects[58] (Methods).

Among the 341 protein mediators identified in pathways linking diabetes to six secondary diseases, 69 showed evidence of potential causal effects on five outcomes: CereD, GD, CAA, HF, and RTID. Similarly, 24 of 131 and 12 of 64 proteins associated with dyslipidemia and hypertension, respectively, had significant MR evidence of causal roles (Supplementary Table 6). Among outcomes, CAA was associated with the largest number of putative causal proteins (38, Bonferroni adjusted[59] $p < 0.05$). Notably, all 38 proteins were identified as significant mediators in the diabetes-CAA pathway, and three were in the dyslipidemia-CAA pathway, whereas no causal mediators were identified for the hypertension-CAA pathway.

Our analysis supported several previously reported causal roles of proteins in disease pathways (Fig. 5b). For example, VCAM1 showed evidence of a potential causal associated with artery diseases (OR = 1.82, $p = 1.36 \times 10^{-7}$), consistent with its established biological function as a key endothelial adhesion molecule that mediates leukocyte recruitment to sites of vascular inflammation, thereby contributing to atherosclerotic plaque formation and development[60]. Similarly, ASGR1, a key protein involved in cholesterol metabolism, showed significant indirect effects in pathways from diabetes, hypertension, and dyslipidemia to multiple types of CVDs. MR analysis further suggested potential causal roles for ASGR1 in CAA (OR = 1.186, $p = 7.06 \times 10^{-6}$) and HF (OR = 1.121, $p = 1.01 \times 10^{-7}$). In addition, PCSK9 inhibitors, a well-established therapeutic target in lipid metabolism[61], was also identified as a potentially causal factor for several CVDs, including angina pectoris (OR = 1.27, $p = 4.44 \times 10^{-13}$). Consistent with this finding, PCSK9 also exhibited significant indirect effects in pathways linking dyslipidemia to three types of CVDs, including HF, VD, and IHD.

Plasma Protein Mediators Improve Prediction of Secondary Disease Risk Among Individuals with Primary Diseases

Using a LASSO-penalized Cox regression model[62], we evaluated the predictive value of identified protein mediators for secondary disease risk among participants with a diagnosed primary disease. Analyses were restricted to disease pathways supported by at least 10 mediating proteins. For each pathway, the top ten proteins ranked by indirect effect size were used as predictors. We used 10-fold nested cross-validation to split the training and test set and to optimize hyperparameters[63]. Predictive performance was assessed using the concordance index[64] (C-index) evaluated on held-out test data. (Methods).

Across disease pathways, models incorporating the top ten mediation-ranked proteins consistently outperformed predictive models based on (i) basic demographic covariates (age, sex, and body mass index), (ii) established lifestyle and clinical biomarkers (including blood pressure, cholesterol levels, smoking, and alcohol consumption), (iii) randomly selected sets of ten proteins (100 repeated samplings), and (iv) the full proteomic panel (Fig. 5c and Supplementary Table 7, Methods). Notably, models incorporating either the entire proteomic profile or random subsets of proteins frequently showed lower predictive accuracy across multiple secondary diseases, indicating that inclusion of non-informative or weakly relevant proteins can dilute predictive signals, underscoring the importance of biologically informed feature selection for improving model performance. To assess the stability of the results, we compared model performance across individual cross-validation folds. Mediation-prioritized proteins achieved the highest predictive accuracy in an average of 70% of folds, with superiority ranging from 50% to 100% across disease pathways (Methods).

We further investigated how the number of proteins included in the models affected prediction accuracy (Fig. 5d). In 10 out of the 12 tested disease pathways, the model using the top 10 plasma proteins outperformed models using only the top five, three, or top one protein. However, in most cases, the predictive performance of models using the top five proteins achieved comparable performance, suggesting that a parsimonious set of high-impact proteins may yield clinically meaningful prediction while improving model interpretability and efficiency.

Discussion

In this study, we systematically investigated the role of plasma proteins in mediating the development from three major cardiometabolic conditions—diabetes, hypertension, and dyslipidemia—to a broad range of downstream diseases using data from 53,030 UK Biobank

participants. We identified 1,461 significant mediation pathways involving 395 unique plasma proteins across 18 secondary conditions, revealing both shared and disease-specific molecular mediators. Mendelian randomization further supported potential causal roles for 84 proteins, highlighting their relevance as candidate therapeutic targets. In addition, mediation-prioritized proteins significantly improved the prediction of secondary disease risk beyond conventional clinical factors and other proteomic features. Functional enrichment analyses indicated that these mediators are predominantly involved in receptor-mediated signaling and molecular interaction pathways. Furthermore, stratification by biomarker-defined disease control and severity revealed additional pathways and mediators, underscoring the importance of disease management in mitigating downstream complications.

Our findings support and extend previous observational studies linking the three primary cardiometabolic diseases--diabetes, hypertension, and dyslipidemia--to a broad range of secondary conditions. Numerous studies have established these disorders as key risk factors for CVDs[3,7,32] and RDs[29,31], with dyslipidemia also implicated in PD[65]. Our results confirmed these well-known associations while also uncovering several underexplored relationships, including a potential link between hypertension and a subtype of PuD (OPDRI). In addition, our analysis confirmed several well-known disease-protein relationships, including those between diabetes and GDF15 and dyslipidemia and PCSK9.

Our results further highlight the central role of plasma proteins as mediators of disease development. Some proteins, such as GDF15 and LGALS4, consistently mediated pathways linking a single primary disease to multiple secondary outcomes, whereas others, such as ASGR1, mediated associations from all three primary diseases to a common outcome, including heart failure. Mendelian randomization analysis further supported potential causal roles for many of these key mediators, highlighting their biological relevance in disease development.

Prediction analysis demonstrated that the mediation-prioritized proteins provided superior predictive performance for secondary disease risk among individuals with primary diseases, outperforming models based on other plasma proteins or conventional clinical risk factors. For example, improving glomerular disease risk prediction by approximately 14% in individuals with hypertension. Notably, 14 out of 69 proteins showed both evidence of causal effects and strong predictive utility. For instance, HAVCR1, which mediated the pathway from hypertension to

valve disorders, also emerged as a key predictor of secondary disease risk (indirect effect: 0.044, OR: 1.01, $p < 1 * 10^{-10}$ from the MR analysis).

The enrichment patterns of mediating proteins cross disease pathways suggest that intercellular signaling represents a central mechanism underlying molecular mediation between cardiometabolic risk factors and their complications. The predominance of pathways related to receptor activity, ligand binding, and cytokine interactions is consistent with a framework in which circulating proteins act as systemic signaling mediators linking dysfunction across organs. In particular, the recurrent enrichment of receptor-mediated signaling pathways in diabetes-associated disease pathways supports the concept that metabolic disease promotes multi-organ injury through coordinated extracellular signaling networks rather than isolated tissue-specific mechanisms. The identification of death receptor signaling, growth factor interactions, and lipid-binding pathways further suggests that inflammatory, apoptotic, and metabolic signaling axes contribute to disease propagation across cardiovascular and renal systems.

Beyond defining primary disease status based on diagnostic history, we also incorporated continuous clinical biomarkers—HbA1c and blood pressure—to capture disease control and severity, enabling stratification of participants into multiple risk categories. HbA1c-based stratification revealed several novel disease pathways not identified using diagnosis alone, including a previously unreported association between poor glycemic control and increased risk of AD. Modeling the nonlinear relationship between HbA1c and AD risk further identified two potential “acceleration points” at 5.8% and 7.9%, beyond which disease risk increased sharply, highlighting the importance of glycemic control in mitigating downstream complications. For hypertension, stratification by blood pressure levels yielded disease pathways largely consistent with those identified using diagnostic history, suggesting that quantitative blood pressure measurements recapitulate established risk structures while providing additional resolution of disease severity. These stratifying findings demonstrate the importance of glucose and blood pressure management in reducing secondary disease risk among individuals with cardiometabolic conditions. We did not stratify dyslipidemia severity using biomarkers such as lipoprotein, because dyslipidemia may involve either elevated or reduced lipid levels, depending on disease subtype or treatment status.

Several limitations of our study should be acknowledged. First, our analysis focused on disease development from three primary cardiometabolic conditions—diabetes, hypertension,

and dyslipidemia—to their downstream complications. Although the analytical framework can be readily extended to other disease pathways, our findings are limited to these cardiometabolic contexts. Second, we examined 18 secondary diseases representing a broad spectrum of clinically relevant outcomes; however, some conditions with very low prevalence ($\leq$ 0.2%) were not included in our study. Third, the indirect effects of plasma proteins were evaluated individually, without accounting for potential joint or interactive effects among proteins, which may underestimate the complexity of the underlying molecular mediation networks. In addition, our analysis assumed linear relationships between protein levels and secondary disease risk, which may oversimplify the potentially nonlinear biological processes underlying disease development. Lastly, the generalizability of our findings may be limited, as the majority of participants in the UK Biobank cohort were white individuals of European ancestry. Further validation in more diverse populations will be necessary to determine whether these mediation pathways are consistent across different genetic and environmental backgrounds.

In conclusion, our findings provide a comprehensive map of plasma protein mediators underlying the development of cardiometabolic disease to secondary outcomes and highlight a parsimonious set of identified protein mediators with potential as biomarkers and therapeutic targets. These results offer new insights into the molecular pathways linking cardiometabolic risk factors to downstream complications and provide a framework for identifying clinically accessible markers to improve risk stratification and guide targeted interventions.

Methods

UK Biobank Plasma Protein Cohort

Plasma proteomic measurements were obtained from the UK Biobank Pharma Proteomics Project (UKB-PPP)[21], which used the Olink Explore 3072 proximity extension assay (PEA) platform, an antibody-based technique. We included 53,030 participants in our study, whose blood samples were collected between March 2006 and September 2010 and measured between April 2021 and February 2022, yielding 2,941 protein analytes corresponding to 2,923 unique proteins. The time point of plasma protein data collection was defined as instance 0. A subset of participants underwent two additional rounds of plasma collection from 2015 to 2019 and again in 2021. However, these profiles covered a smaller panel of proteins (1,173), with only 506 proteins overlapping with those collected at instance 0. Due to this limited overlap, these

later profiles were not included in the primary analyses. In addition, GLIPR1 was excluded due to insufficient data availability, resulting in 2,922 plasma proteins included in the final analyses.

Primary Diseases and Secondary Diseases

The primary diseases analyzed in this study included diabetes, hypertension, and dyslipidemia, which are well-known risk factors for multiple subsequent diseases. In this study, we examined secondary diseases across four major disease categories, including cardiovascular, pulmonary, renal, and neurological disease (Fig. 1). After excluding disease subtypes with very low prevalence (< 0.1%), a total of 18 secondary diseases were included in the analysis. Disease categorization followed the original UK Biobank disease classification framework and was further reviewed by clinical expertise. Detailed categorization, corresponding ICD-10 codes, and disease prevalence are provided in Supplementary Table 3.

Primary disease status was defined as a binary variable, indicating whether participants were diagnosed with a primary disease within a window from two years before to one year after the plasma protein collection (instance 0). Individuals diagnosed more than two years before instance 0 were excluded to ensure that primary disease status closely reflected the time point of plasma protein collection. The number of cases for each primary disease ranged from 1,053 (2.0%) to 3,012 (5.7%).

Participants diagnosed with secondary diseases prior to instance 0 or before the onset of the primary disease were excluded, as the analysis focused on the future risk of secondary disease development. Follow-up time and outcome information were obtained from the UK Biobank longitudinal health records. For each outcome, follow-up began at instance 0 and continued until the earliest occurrence of the outcome event or the end of follow-up (September 1, 2023).

Covariates

The covariates considered in this study included participants' age, sex, race, body mass index (BMI), smoking status, alcohol consumption status, diastolic blood pressure (DBP), systolic blood pressure (SBP), and glycated hemoglobin (HbA1c). These variables are well-established risk factors and potential confounders for cardiometabolic and chronic diseases[66–68], including cardiovascular, renal, neurological, and pulmonary diseases. Information on smoking

status and alcohol consumption was collected through questionnaires administered at baseline (instance 0). Smoking status was categorized into three levels (never, past, and current), and alcohol consumption was categorized into five levels (never, monthly or less, 2–4 times per month, 2–3 times per week, and ≥4 times per week). Measurements of SBP, DBP, and HbA1c were also obtained at baseline. To avoid over-adjustment, covariates that directly overlapped with the definitions of primary diseases were excluded from certain models. Specifically, SBP and DBP were excluded when hypertension was the primary disease, whereas HbA1c was retained to account for potential confounding by diabetes. Conversely, HbA1c was excluded in models where diabetes was the primary disease, while SBP and DBP were retained.

Stratification by Blood Pressure and Glycated Hemoglobin

In addition to defining primary disease status based on diagnostic history, we further stratified participants according to SBP, DBP, and HbA1c levels. Unlike historical disease data, blood pressure and glycated hemoglobin measurements were obtained on the same day as the plasma protein data collection (instance 0). However, these physiological markers may be more sensitive to short-term environmental or individual factors, such as medication use and recent physical activity. Therefore, blood pressure and glycated hemoglobin levels were used as complementary indicators of disease severity and control alongside diagnostic history.

According to the 2023 European Society of Hypertension (ESH) guidelines[45], participants were classified into four blood pressure categories: normal (< 140/90 mmHg), hypertension grade 1 (140–159/90–99 mmHg), hypertension grade 2 (160–179/100–109 mmHg), and hypertension grade 3 (≥ 180/110 mmHg). Similarly, participants were classified based on HbA1c levels[39] into normal (HbA1c < 5.7%), prediabetes (5.7% ≤ HbA1c < 6.4%), and severe-diabetes (HbA1c ≥ 6.4%).

Using these categorizations, we examined whether participants with elevated blood pressure (grades 1 to 3) or impaired glycemic control (prediabetes and severe-diabetes) exhibited an increased likelihood of developing secondary diseases compared to those with normal levels. Mediation analysis was then performed to assess the role of plasma proteins in mediating associations between disease severity or control and secondary disease risk.

Mediation Analysis

Under the mediation analysis framework[21], we defined the onset of primary disease as the exposure variable, baseline expression levels of 2,922 plasma proteins as potential mediators, and the occurrence and timing of secondary diseases as outcome variables. We first evaluated the association between primary disease status and secondary disease risk using a Cox proportional hazards model[69],

$$\log\left(\frac{H(t)}{H(t_0)}\right) = cX + z_1^T\Omega \ \#(1)$$

where $H(t)$ denotes the hazard function at time t, t_0 is the time at the instance 0, c is the log relative hazard (defined as the total effect), X is the binary primary disease status (diagnosed vs. not diagnosed), Ω is the vector of covariates, and z_1 denotes the corresponding coefficients. Covariates were described in the Covariate section. Secondary diseases showing statistically significant associations with primary diseases ($p < 0.05$) were advanced to subsequent analyses.

Next, we assessed the associations between plasma protein levels and primary disease status using linear regression models with the same covariate set,

$$M_i = aX + z_2^T\Omega \ \#(2)$$

where, M_i is the expression level of protein i, a represents the protein expression change associated with primary disease status, and z_2 represents the coefficients for covariates. Proteins were considered significantly associated with the primary disease if the BH adjusted p-values were < 0.05 and were subsequently advanced to further analyses.

Finally, we examined the potential mediating role of plasma proteins using a Cox model incorporating both exposure and mediator variables,

$$\log\left(\frac{H(t)}{H(t_0)}\right) = c'X + bM + z_3^T\Omega. \#(3)$$

where c' represents the direct effect of the primary disease, b represents the effect of one plasma protein, and z_3 epresents covariate effects. The indirect effect of each protein was estimated using the product-of-coefficients method[21] ($a \cdot b$) and its statistical significance was evaluated via 1,000 bootstrap resamples[70]. Resulting p-values were adjusted suing the BH procedure. This analytical pipeline was applied for each disease pathway.

Plasma proteins were defined as significant mediators if they met all the following criteria: (i) a statistically significant indirect effect after multiple testing correction, (ii) an absolute value of the indirect effect ≥ 0.01, and (iii) a proportional indirect effect (PIE), define as the ratio of the

indirect to the total effect, of at least 10%. We further defined key protein mediators as those that: (i) consistently mediated pathways from a single primary disease to multiple secondary diseases; (ii) exhibited strong mediation effects within individual disease pathways; or (iii) mediated pathways that linked all three primary diseases to a common secondary disease.

Mendelian Randomization

To evaluate the potential causal effects of identified plasma mediators on their corresponding secondary diseases, we performed two-sample Mendelian Randomization (MR)[71]. We used *cis*-pQTLs as instrument variable (within 1 Mb of the transcription start site) from the UKB-PPP at genome-wide significance ($p < 5 \times 10^{-11}$)[23]. Independent instruments were selected through linkage disequilibrium (LD) clumping ($r^2 < 0.01$) using the European reference panel, discarding instruments with F-statistics<10 to mitigate weak instrument bias. Genetic associations of the selected variants with secondary diseases were extracted from publicly available genome-wide association study (GWAS) summary statistics from the IEU OpenGWAS database[56,72], restricted to individuals of European ancestry. Harmonization procedure was performed to align effect alleles across exposure (pQTLs) and disease outcome GWAS datasets, and palindromic variants with ambiguous strand orientation were excluded.

Twelve MR methods, including inverse-variance weighted (IVW) method (fixed effect, multiplicative random effects with/without corrected standard error under dispersion), MR-Egger regression, Simple median, Weighted median, Penalised weighted median, Simple mode, Simple mode (NOME), Weighted mode, Weighted mode (NOME), were applied[58]. For single-variant instruments, the Wald ratio was used. Heterogeneity across instruments was assessed using Cochran's Q statistic and the I^2 metric, and horizontal pleiotropy was evaluated using the MR–Egger intercept. An association was considered to show a potential causal relationship if the BH adjusted p-value was <0.05 in at least one MR method. All analyses were conducted using the TwoSampleMR package in R[56].

Predicting the Future Risk of Secondary Diseases

To evaluate the predictive power of identified plasma protein mediators, we applied a LASSO-penalized Cox survival model[62] using the top ten proteins with the largest indirect effects as predictors. For each primary–secondary disease pathway, an individual model was

trained to predict the future risk of the corresponding secondary disease among participants diagnosed with the primary disease. We used the nested cross-validation[63] to train the model and evaluate its prediction performance. In the outer loop, participants were randomly partitioned into ten equal folds. In each iteration, nine folds were used for model training and the remaining fold for testing, ensuring that predictions were evaluated exclusively from models trained on non-overlapping samples. This process was repeated across all folds to obtain out-of-sample predictions for the entire cohort. Within each training set, hyperparameters were optimized via inner ten-fold cross-validation. Model performance was evaluated using the concordance index (C-index)[64].

The prediction performance of this mediation-informed model was compared with four alternative approaches: (1) a LASSO survival model including only basic demographic variables (age, sex, BMI, and race); (2) a LASSO survival model incorporating conventional clinical and lifestyle risk factors, including age, sex, BMI, blood pressure, HbA1c level, smoking and drinking statues; (3) a LASSO survival model using ten randomly selected plasma proteins, with the random selection repeated 100 times to obtain the average C-index; (4) an XGBoost survival model[73,74] using all 2,922 plasma proteins as predictors. Due to the substantial proportion of missing values when including all plasma proteins (10.3% missingness), LASSO-based models were not suitable. Therefore, we employed an XGBoost-based model, which is more robust to missing data. For comparability, all models used identical training–test splits defined by a fixed random seed.

Sensitivity Analysis

To evaluate the robustness of proteins identified through mediation analysis, we performed a sensitivity analysis with additional adjustment for 14 clinical and biochemical covariates. These included markers of hepatic function (alanine aminotransferase and aspartate aminotransferase), hematologic parameters (mean corpuscular volume, hemoglobin concentration, and platelet count), renal function indicators (cystatin C, creatinine, and urea), metabolic markers (glucose, triglycerides, low-density lipoprotein cholesterol, total cholesterol, and albumin), as well as socioeconomic status measured by the Townsend deprivation index (TDI). We then compared the top five proteins (ranked by indirect effect) identified for each

disease pathway between the primary covariate model and the extended model including these additional covariates, in terms of Jaccard index[38].

Biological Function Enrichment Analysis

Gene Ontology Biological Process (GO:BP) over-representation analysis was performed for the identified plasma protein mediators using the R package clusterProfiler (v4.16.0)[54]. Plasma proteins were mapped to human gene identifiers using org.Hs.eg.db (v3.21.0) annotation database, retaining unique mapped genes and excluding unmapped entries[75]. Enrichment analysis was conducted with the function enrichGO(·). The background gene set was defined as all 2,922 proteins quantified in the proteomic platform, to control for platform coverage. GO terms were considered significantly enriched if the BH adjusted p-values were < 0.05. To reduce redundancy among enriched GO terms, semantic similarity analysis was performed using the simplify(·) function in the clusterProfiler package. GO terms with a pairwise semantic similarity greater than 0.7 were considered redundant, and the term with the lowest BH adjusted p-value was retained as the representative term.

Data availability

All UK Biobank data (plasma proteome profile and longitudinal clinical data) are accessible via application to the UK Biobank. PQTL summary statistics from the UKB-PPP can be accessed at https://metabolomips.org/ukbbpgwas.

Code availability

All analyses and visualization were implemented in R (v4.4.0). Scripts are available upon reasonable request.

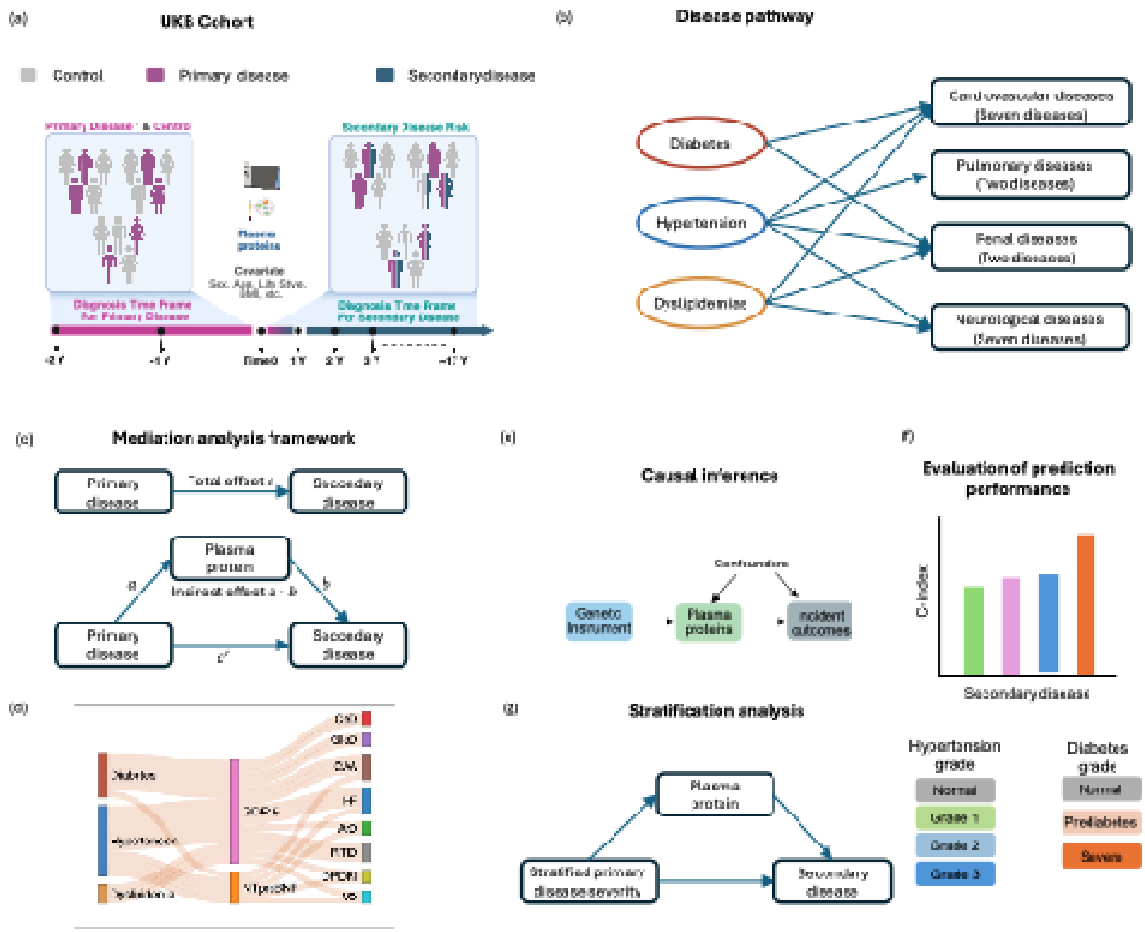


Fig. 1. **Overview of the study design and analytical framework.** (a) Study cohort comprising 53,030 UK Biobank participants with plasma proteomic profiles and longitudinal clinical data. We used this cohort to investigate the mediating roles of plasma proteins linking primary diseases to secondary outcomes. (b) Disease pathways examined from three major cardiometabolic diseases—diabetes, hypertension, and dyslipidemia—to 18 secondary diseases across four clinical domains. (c) Mediation analysis framework, with primary disease status as the exposure, plasma protein expression as the mediator, and secondary disease risk as the outcome. (d) Examples of identified protein mediators linking primary diseases to secondary diseases. For instance, GDF15 and NT-proBNP act as key mediators bridging multiple disease pathways. (e) Causal inference framework using Mendelian randomization to evaluate the potential causal effects of identified protein mediators on secondary disease risk, with genetic variants as instrumental variables. (f) Illustration of the evaluation of prediction performance (C-index) using top-ranked protein mediators for secondary disease risk. (g) Stratification of

primary diseases by severity and control (e.g., HbA1c and blood pressure), revealing additional protein-mediated disease pathways. (a) was created with BioRender.com.

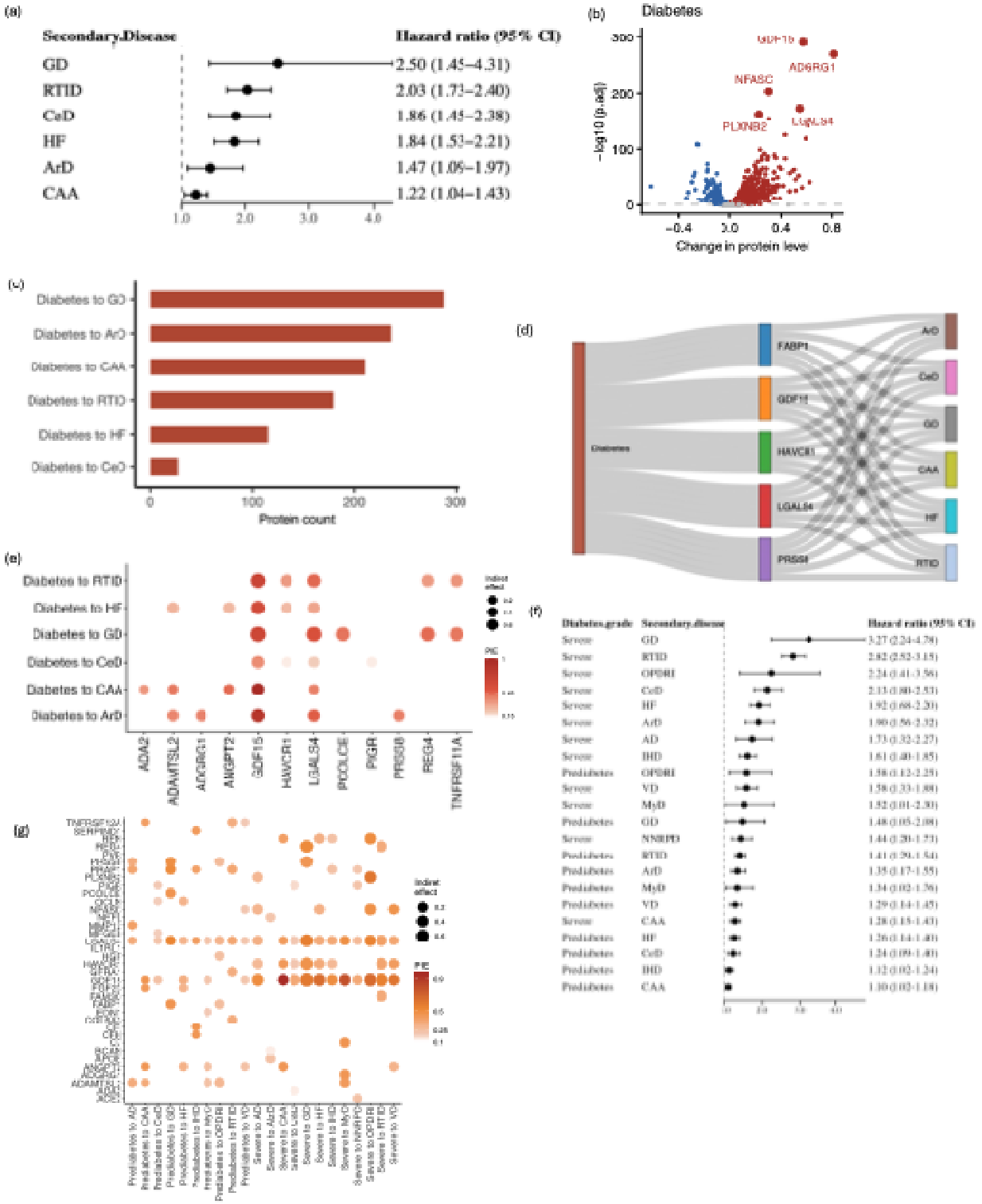


Fig. 2. Results of diabetes-associated plasma protein mediating pathways. (a) Forest plot showing hazard ratios (HRs) with 95% confidence intervals (CIs) for secondary diseases associated with diagnosed diabetes, estimated using Cox proportional hazards models (only associations with $p < 0.05$ are shown). (b) Volcano plot showing changes in plasma protein expression associated with diabetes status and corresponding significance levels from linear models; the top five most significant proteins are highlighted. (c) Number of identified plasma protein mediators across diabetes-associated disease pathways. (d) Sankey diagram showing the top five protein mediators linking diabetes to the greatest number of secondary diseases. (e) Dot plot of the top five protein mediators ranked by indirect effect size in diabetes-associated pathways. Dot size represents the indirect effect, and dot color represents the proportional indirect effect (PIE). (f) Forest plot showing HRs with 95% CIs for secondary disease risk across three diabetes-control groups (only associations with $p < 0.05$ are shown). (g) Dot plot showing the top five protein mediators across diabetes-control groups identified through stratification analysis.

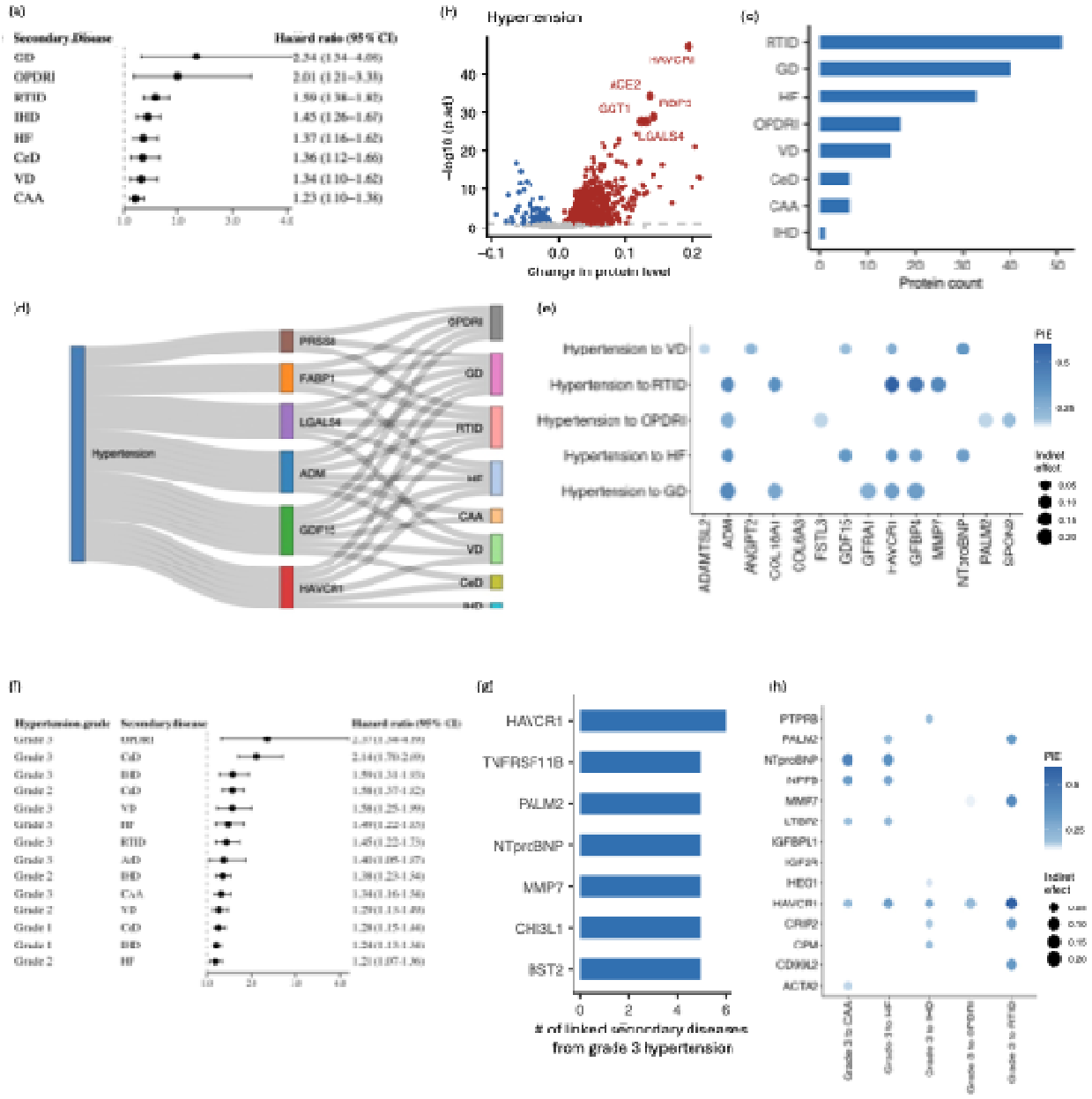


Fig. 3. Results of hypertension-associated plasma protein mediating pathways. (a) Forest plot showing hazard ratios (HRs) with 95% confidence intervals (CIs) for secondary diseases associated with diagnosed hypertension, estimated using Cox proportional hazards models (only associations with $p < 0.05$ are shown). (b) Volcano plot showing changes in plasma protein expression associated with hypertension status and corresponding significance levels from linear models; the top five most significant proteins are highlighted. (c) Number of identified plasma protein mediators across hypertension-associated disease pathways. (d) Sankey diagram showing

the top five protein mediators linking hypertension to the greatest number of secondary diseases. (e) Dot plot of the top five protein mediators ranked by indirect effect size in hypertension-associated pathways. Dot size represents the indirect effect, and dot color represents the proportional indirect effect (PIE). (f) Forest plot showing HRs with 95% CIs for secondary disease risk across three hypertension severity groups (only associations with $p < 0.05$ are shown). (g) Top protein mediators most frequently linking grade 3 hypertension to secondary diseases. (h) Dot plot of the top five protein mediators ranked by indirect effect in pathways linking grade 3 hypertension to secondary diseases.

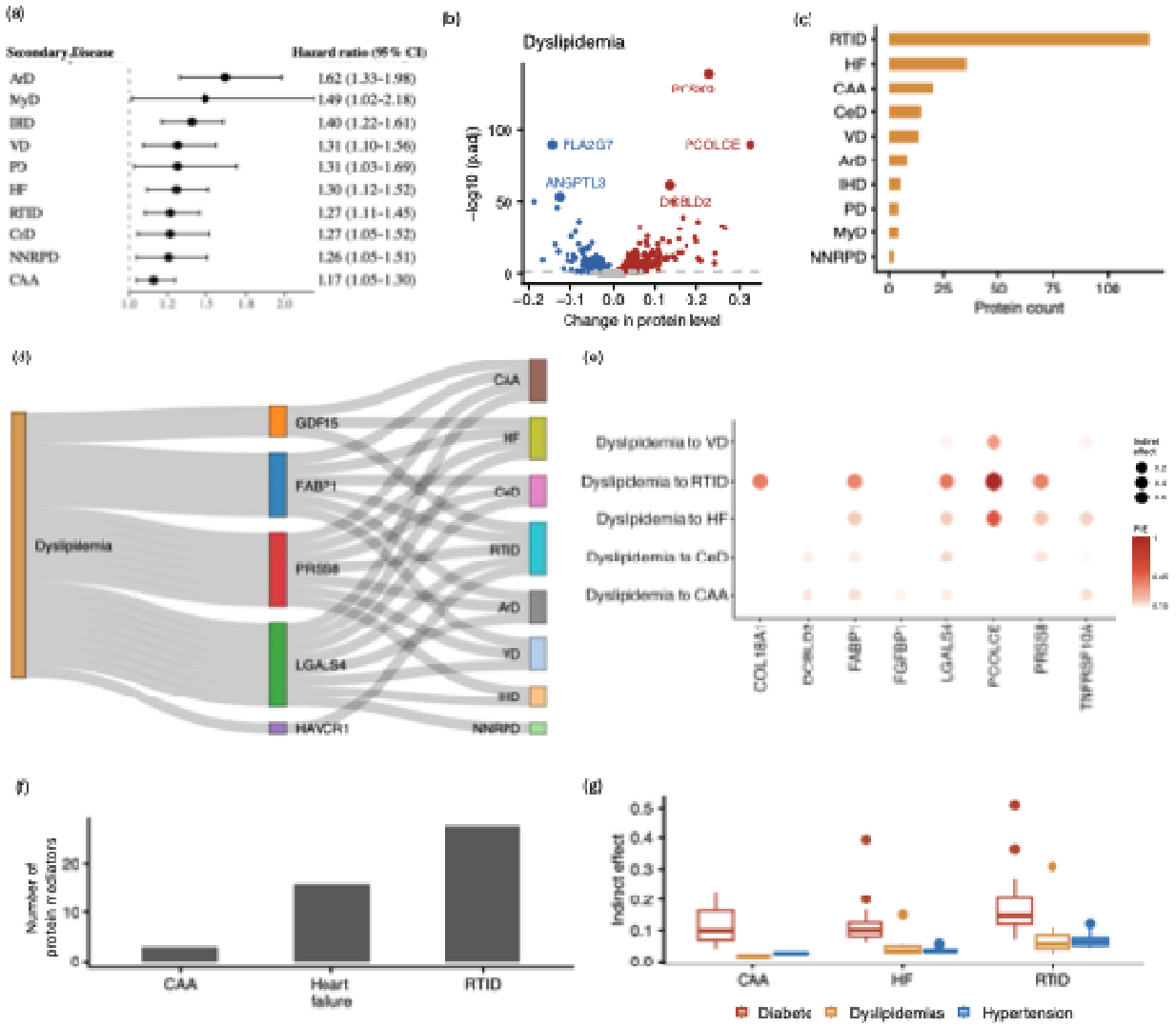


Fig. 4. Results of dyslipidemia-associated plasma protein mediating pathways. (a) Forest plot showing hazard ratios (HRs) with 95% confidence intervals (CIs) for secondary diseases associated with diagnosed dyslipidemia, estimated using Cox proportional hazards models (only associations with $p < 0.05$ are shown). (b) Volcano plot showing changes in plasma protein expression associated with dyslipidemia status and corresponding significance levels from linear models; the top five most significant proteins are highlighted. (c) Number of identified plasma protein mediators across dyslipidemia-associated disease pathways. (d) Sankey diagram showing the top five protein mediators linking dyslipidemia to the greatest number of secondary diseases. (e) Dot plot of the top five protein mediators ranked by indirect effect size in dyslipidemia-associated pathways. Dot size represents the indirect effect, and dot color represents the proportional indirect effect (PIE). (f) Number of shared protein mediators linking all three

primary diseases to CAA, HF, and RTID. (g) Boxplot showing the distribution of indirect effects for these shared mediators across the three primary diseases.

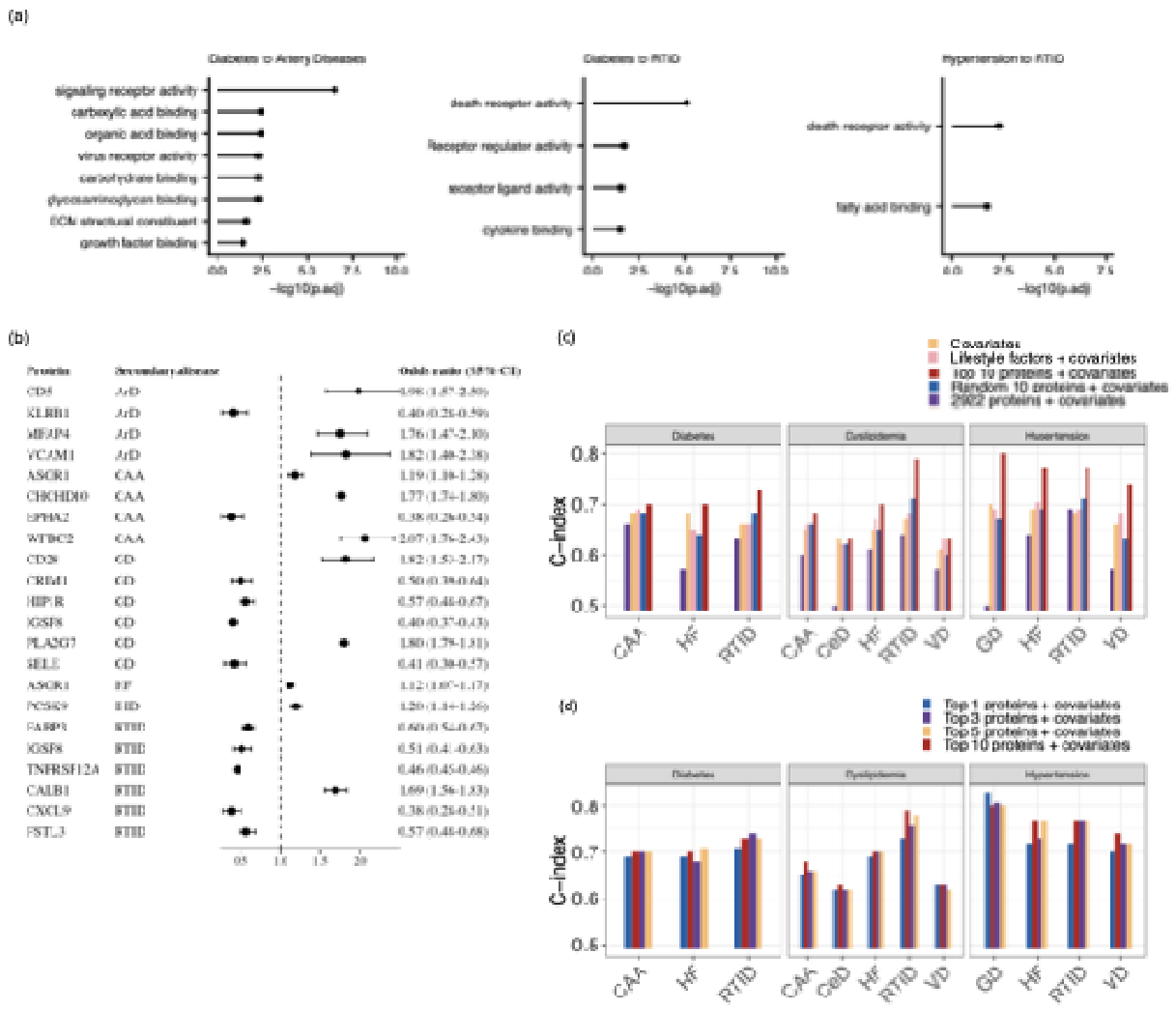


Fig. 5. Biological function annotation of plasma protein mediators and evaluation of their causal and predictive effects. (a) Gene Ontology (GO) enrichment analysis of identified protein mediators across three representative pathways (full results are provided in the Supplementary Table 12). (b) Forest plot showing odds ratios (ORs) with 95% CIs for secondary disease risk associated with protein mediators, as estimated by Mendelian randomization. (c) Comparison of prediction performance (C-index) for secondary disease risk across three primary diseases using different models: (i) all plasma proteins with covariates (age, sex, and BMI), (ii) covariates only, (iii) lifestyle factors and covariates, (iv) randomly selected sets of 10 plasma proteins with covariates (average performance over 100 replicates), and (v) the top 10 mediation-ranked protein mediators with covariates. (d) Comparison of prediction performance (C-index) using the top 1, 3, 5, and 10 mediation-ranked protein mediators with covariates.

Table. 1. Secondary Disease Names and Corresponding Abbreviations

Organ system	Disease name	Abbreviations
Pulmonary diseases	Chronic lower respiratory diseases	CLRD
Pulmonary diseases	Other respiratory diseases principally affecting the interstitium	OPDRI
Renal diseases	Renal tubulo-interstitial diseases	RTID
Renal diseases	Glomerular diseases	GD
Neurological diseases	Systemic atrophies primarily affecting the central nervous system	SACNS
Neurological diseases	Parkison's disease	PD
Neurological diseases	Alzheimer's disease	ad
Neurological diseases	Multiple sclerosis	MS
Neurological diseases	Nerve nerve root and plexus disorders	NNRPD
Neurological diseases	Diseases of myoneural junction and muscle	MNJM
Cardiovascular diseases	Ischaemic heart diseases	IHD
Cardiovascular diseases	Valve disorder	VD
Cardiovascular diseases	Myocardial Disorders	MyD
Cardiovascular diseases	Cardiac arrest and arrhythmias	CAA
Cardiovascular diseases	Heart failure	HF
Cardiovascular diseases	Cerebrovascular diseases	CeD
Cardiovascular diseases	Artery disease	ArD

References

1. Ghazwani, M. *et al.* Prevalence of Dyslipidemia and Its Determinants Among the Adult Population of the Jazan Region. *Int. J. Gen. Med.* **Volume 16**, 4215–4226 (2023).
2. Kannel, W. B. Hypertension: relationship with other risk factors. *Drugs* **31**, 1–11 (1986).
3. Sowers, J. R., Epstein, M. & Frohlich, E. D. Diabetes, Hypertension, and Cardiovascular Disease. *Hypertension* **37**, 1053–1059 (2001).
4. Rosano, G. M. C., Vitale, C. & Seferovic, P. Heart failure in patients with diabetes mellitus. *Card. Fail. Rev.* **3**, 52 (2017).
5. Almdal, T., Scharling, H., Jensen, J. S. & Vestergaard, H. The independent effect of type 2 diabetes mellitus on ischemic heart disease, stroke, and death: a population-based study of 13 000 men and women with 20 years of follow-up. *Arch. Intern. Med.* **164**, 1422–1426 (2004).
6. Matheus, A. S. de M. *et al.* Impact of diabetes on cardiovascular disease: an update. *Int. J. Hypertens.* **2013**, 653789 (2013).
7. Resnick, H. E. & Howard, B. V. Diabetes and cardiovascular disease. *Annu. Rev. Med.* **53**, 245–267 (2002).
8. Whelton, P. K. & Klag, M. J. Hypertension as a risk factor for renal disease. Review of clinical and epidemiological evidence. *Hypertension* **13**, I19 (1989).
9. Freitas Lima, L. C. *et al.* Adipokines, diabetes and atherosclerosis: an inflammatory association. *Front. Physiol.* **6**, (2015).
10. Savoia, C. & Schiffrin, E. L. Vascular inflammation in hypertension and diabetes: molecular mechanisms and therapeutic interventions. *Clin. Sci.* **112**, 375–384 (2007).
11. Granger, D. N., Vowinkel, T. & Petnehazy, T. Modulation of the Inflammatory Response in Cardiovascular Disease. *Hypertension* **43**, 924–931 (2004).
12. Assarsson, E. *et al.* Homogenous 96-Plex PEA Immunoassay Exhibiting High Sensitivity, Specificity, and Excellent Scalability. *PLoS One* **9**, e95192 (2014).
13. Gold, L. *et al.* Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery. *PLoS One* **5**, e15004 (2010).
14. Oh, H. S.-H. *et al.* Plasma proteomics links brain and immune system aging with healthspan and longevity. *Nat. Med.* https://doi.org/10.1038/s41591-025-03798-1 (2025) doi:10.1038/s41591-025-03798-1.
15. Oh, H. S.-H. *et al.* Organ aging signatures in the plasma proteome track health and disease. *Nature* **624**, 164–172 (2023).
16. Carrasco-Zanini, J. *et al.* Proteomic signatures improve risk prediction for common and rare diseases. *Nat. Med.* **30**, 2489–2498 (2024).
17. Gadd, D. A. *et al.* Blood protein assessment of leading incident diseases and mortality in the UK Biobank. *Nat. Aging* **4**, 939–948 (2024).
18. You, J. *et al.* Plasma proteomic profiles predict individual future health risk. *Nat. Commun.* **14**, 7817 (2023).
19. qian, hanyu *et al.* Uncovering the Molecular Landscape of Physical Activity: Proteomic Insights from the UK Biobank. Preprint at https://doi.org/10.1101/2025.06.27.25330418 (2025).
20. Deng, Y.-T. *et al.* Atlas of the plasma proteome in health and disease in 53,026 adults. *Cell* https://doi.org/10.1016/j.cell.2024.10.045 (2024) doi:10.1016/j.cell.2024.10.045.
21. MacKinnon, D. P., Fairchild, A. J. & Fritz, M. S. Mediation Analysis. *Annu. Rev. Psychol.* **58**, 593–614 (2007).

22. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
23. Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
24. Wrighten, S. A., Piroli, G. G., Grillo, C. A. & Reagan, L. P. A look inside the diabetic brain: Contributors to diabetes-induced brain aging. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* **1792**, 444–453 (2009).
25. Seeger, W. *et al.* Pulmonary hypertension in chronic lung diseases. *J. Am. Coll. Cardiol.* **62**, D109–D116 (2013).
26. Robins, S. J., Lyass, A., Zachariah, J. P., Massaro, J. M. & Vasan, R. S. Insulin resistance and the relationship of a dyslipidemia to coronary heart disease: the Framingham Heart Study. *Arterioscler. Thromb. Vasc. Biol.* **31**, 1208–1214 (2011).
27. Barri, Y. M. Hypertension and kidney disease: a deadly connection. *Curr. Hypertens. Rep.* **10**, 39–45 (2008).
28. Phillips, S. J. & Whisnant, J. P. Hypertension and the brain. *Arch. Intern. Med.* **152**, 938–945 (1992).
29. Wanner, C. & Quaschning, T. Dyslipidemia and renal disease: pathogenesis and clinical consequences. *Curr. Opin. Nephrol. Hypertens.* **10**, 195–201 (2001).
30. Ghaderian, S. B., Hayati, F., Shayanpour, S. & Mousavi, S. S. B. Diabetes and end-stage renal disease; a review article on new concepts. *J. Renal Inj. Prev.* **4**, 28 (2015).
31. Bartges, J. W., Willis, A. M. & Polzin, D. J. Hypertension and Renal Disease. *Veterinary Clinics of North America: Small Animal Practice* **26**, 1331–1345 (1996).
32. Hedayatnia, M. *et al.* Dyslipidemia and cardiovascular disease risk among the MASHAD study population. *Lipids Health Dis.* **19**, 42 (2020).
33. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**, 289–300 (1995).
34. Adela, R. & Banerjee, S. K. GDF□15 as a target and biomarker for diabetes and cardiovascular diseases: a translational prospective. *J. Diabetes Res.* **2015**, 490842 (2015).
35. Wang, D. *et al.* GDF15: emerging biology and therapeutic applications for obesity and cardiometabolic disease. *Nat. Rev. Endocrinol.* **17**, 592–607 (2021).
36. Li, J. *et al.* Overview of growth differentiation factor 15 (GDF15) in metabolic diseases. *Biomedicine & Pharmacotherapy* **176**, 116809 (2024).
37. Dieden, A. *et al.* Galectin-4 is associated with diabetes and obesity in a heart failure population. *Sci. Rep.* **13**, 20285 (2023).
38. Fletcher, S. & Islam, M. Z. Comparing sets of patterns with the Jaccard index. *Australasian Journal of Information Systems* **22**, (2018).
39. ElSayed, N. A. *et al.* 2. Diagnosis and Classification of Diabetes: Standards of Care in Diabetes—2025. *Diabetes Care* **48**, S27–S49 (2025).
40. Baglietto-Vargas, D., Shi, J., Yaeger, D. M., Ager, R. & LaFerla, F. M. Diabetes and Alzheimer's disease crosstalk. *Neurosci. Biobehav. Rev.* **64**, 272–287 (2016).
41. Fernández-Calle, R. *et al.* APOE in the bullseye of neurodegenerative diseases: impact of the APOE genotype in Alzheimer's disease pathology and brain diseases. *Mol. Neurodegener.* **17**, 62 (2022).

42. Chaiyagot, N. *et al.* Exploring Kidney Injury Molecule-1 and HAVCR1 Polymorphisms as Predictive Biomarkers in Chronic Kidney Disease. *Kidney Diseases* **11**, 342–355 (2025).
43. Ishimitsu, T. *et al.* Genomic Structure of Human Adrenomedullin Gene. *Biochem. Biophys. Res. Commun.* **203**, 631–639 (1994).
44. Wong, H. K., Cheung, T. T. & Cheung, B. M. Y. Adrenomedullin and cardiovascular diseases. *JRSM Cardiovasc. Dis.* **1**, 1–7 (2012).
45. Mancia, G. *et al.* 2023 ESH Guidelines for the management of arterial hypertension The Task Force for the management of arterial hypertension of the European Society of Hypertension. *J. Hypertens.* **41**, 1874–2071 (2023).
46. Panagopoulou, V. *et al.* NTproBNP: an important biomarker in cardiac diseases. *Curr. Top. Med. Chem.* **13**, 82–94 (2013).
47. Seidah, N. G. PCSK9 as a therapeutic target of dyslipidemia. *Expert Opin. Ther. Targets* **13**, 19–28 (2009).
48. Wang, G., Bonkovsky, H. L., de Lemos, A. & Burczynski, F. J. Recent insights into the biological functions of liver fatty acid binding protein 1. *J. Lipid Res.* **56**, 2238–2247 (2015).
49. Sawicki, L. R. *et al.* FABP1 knockdown in human enterocytes impairs proliferation and alters lipid metabolism. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* **1862**, 1587–1594 (2017).
50. Xiao, L. *et al.* Echinocystic acid prevents obesity and fatty liver via interacting with FABP1. *Phytotherapy Research* **37**, 3617–3630 (2023).
51. Lely, A. T., Hamming, I., van Goor, H. & Navis, G. J. Renal ACE2 expression in human kidney disease. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland* **204**, 587–593 (2004).
52. Aleksander, S. A. *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**, (2023).
53. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
54. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
55. Sanderson, E. *et al.* Mendelian randomization. *Nature Reviews Methods Primers* **2**, 6 (2022).
56. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, (2018).
57. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* **13**, e1007081 (2017).
58. Wu, C. *et al.* Large-Scale Identification of Novel Protein Biomarkers and Therapeutic Targets in Heart and Brain Disease. Preprint at https://doi.org/10.64898/2026.01.26.26344874 (2026).
59. Etymologia: Bonferroni Correction. *Emerg. Infect. Dis.* **21**, (2015).
60. Ley, K. & Huo, Y. VCAM-1 is critical in atherosclerosis. *Journal of Clinical Investigation* **107**, 1209–1210 (2001).
61. Sobati, S. *et al.* PCSK9: A Key Target for the Treatment of Cardiovascular Disease (CVD). *Advanced Pharmaceutical Bulletin* vol. 10 502–511 Preprint at https://doi.org/10.34172/apb.2020.062 (2020).

62. Tibshirani, R. The lasso method for variable selection in the Cox model. *Stat. Med.* **16**, 385–395 (1997).
63. Hastie, T., Tibshirani, R., Friedman, J. H. & Friedman, J. H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. vol. 2 (Springer, 2009).
64. Steck, H., Krishnapuram, B., Dehing-Oberije, C., Lambin, P. & Raykar, V. C. On ranking in survival analysis: Bounds on the concordance index. *Adv. Neural Inf. Process. Syst.* **20**, (2007).
65. Fu, X. *et al.* A systematic review and meta-analysis of serum cholesterol and triglyceride levels in patients with Parkinson's disease. *Lipids Health Dis.* **19**, 97 (2020).
66. Kearns, K., Dee, A., Fitzgerald, A. P., Doherty, E. & Perry, I. J. Chronic disease burden associated with overweight and obesity in Ireland: the effects of a small BMI reduction at population level. *BMC Public Health* **14**, 143 (2014).
67. Shield, K. Chronic diseases and conditions related to alcohol use. *Alcohol Res.* **35**, 155 (2013).
68. Campos, T. da S. *et al.* Cigarette smoking among patients with chronic diseases. *Int. J. Cardiol.* **174**, 808–810 (2014).
69. Abd ElHafeez, S. *et al.* Methods to Analyze Time□to□Event Data: The Cox Regression Analysis. *Oxid. Med. Cell. Longev.* **2021**, (2021).
70. Tibshirani, R. J. & Efron, B. An introduction to the bootstrap. *Monographs on statistics and applied probability* **57**, (1993).
71. Richmond, R. C. & Davey Smith, G. Mendelian Randomization: Concepts and Scope. *Cold Spring Harb. Perspect. Med.* **12**, a040501 (2022).
72. Elsworth, B. *et al.* The MRC IEU OpenGWAS data infrastructure. Preprint at https://doi.org/10.1101/2020.08.10.244293 (2020).
73. Chen, T. *et al.* Xgboost: extreme gradient boosting. *R package version 0.4-2* **1**, 1–4 (2015).
74. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* vols 13-17-August-2016 785–794 (Association for Computing Machinery, 2016).
75. Carlson, M., Falcon, S., Pages, H. & Li, N. org. Hs. eg. db: Genome wide annotation for Human. *R package version* **3**, 3 (2019).

Acknowledgements

M.W., and H.Q., was supported by M.W.'s startup fund from the Institute for Heart and Brain Health. This work was supported by grants the NIH R35 HL155318 to A.R. and the American Heart Association (AHA) grants 23MERIT1038415 to A.R. and SFRN (URLs: https://doi.org/10.58275/ AHA.24SFRNPCN1284382.pc.gr.194135 and https://doi.org/10.58275/ AHA.24SFRNCCN1276092.pc.gr.194131) to A.R. and M.W.

Author contributions

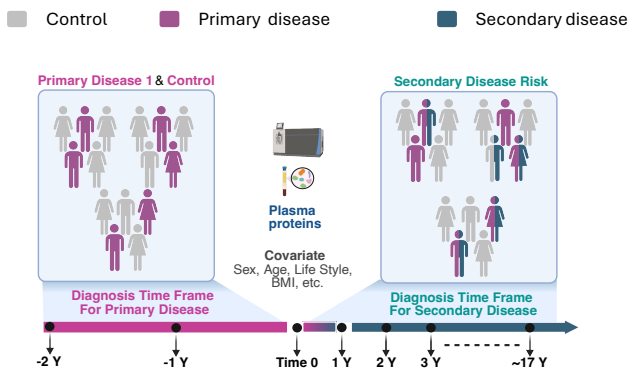
M.W. conceived the study, supervised the research, and revised the manuscript. A.R. revised the manuscript and supervised the research. H.Q. performed the analyses and wrote the manuscript. C.W. supported pathway and biological process analyses and revised the manuscript. B.L. revised the manuscript and figure preparation.

Competing interests

The authors declare no competing interests.

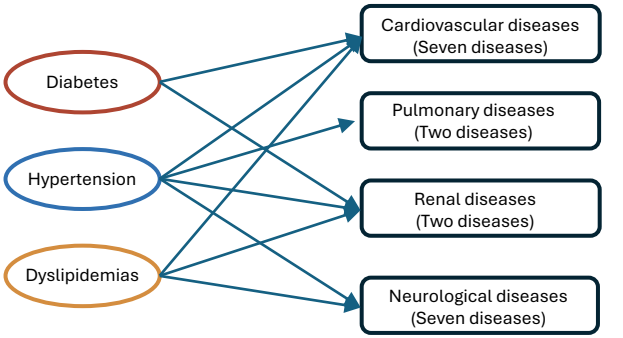
(a)

UKB Cohort



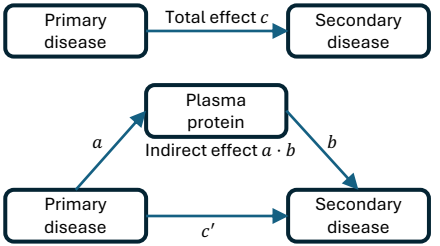
(b)

Disease pathway

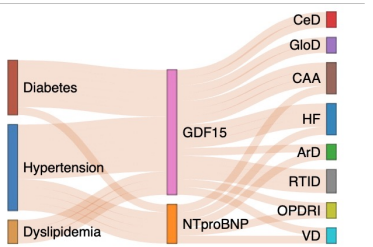


(c)

Mediation analysis framework

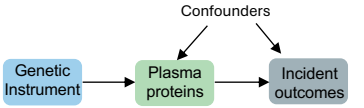


(d)



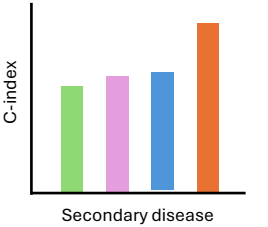
(e)

Causal inference



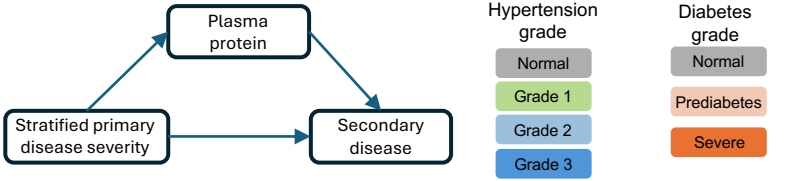
(f)

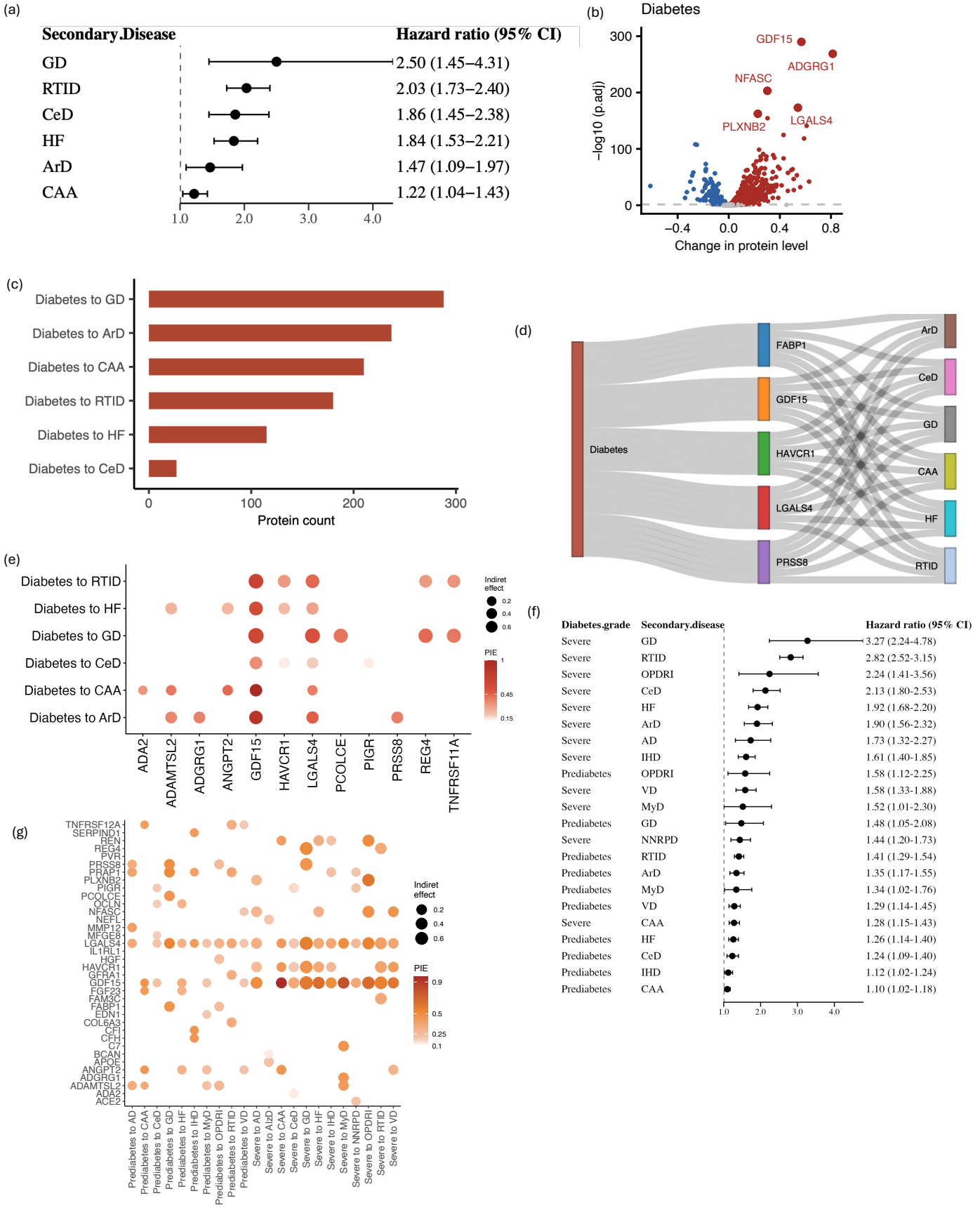
Evaluation of prediction performance

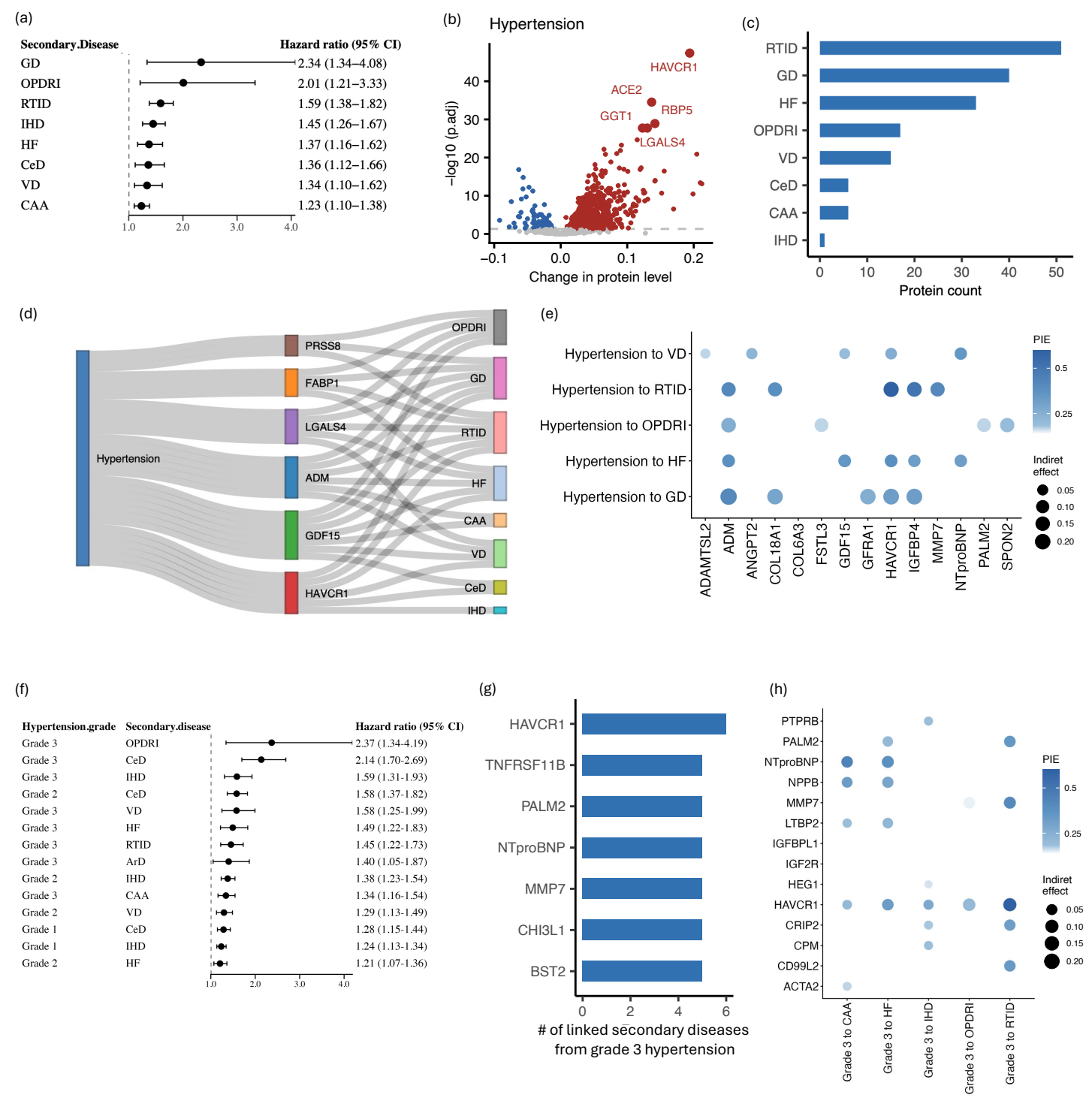


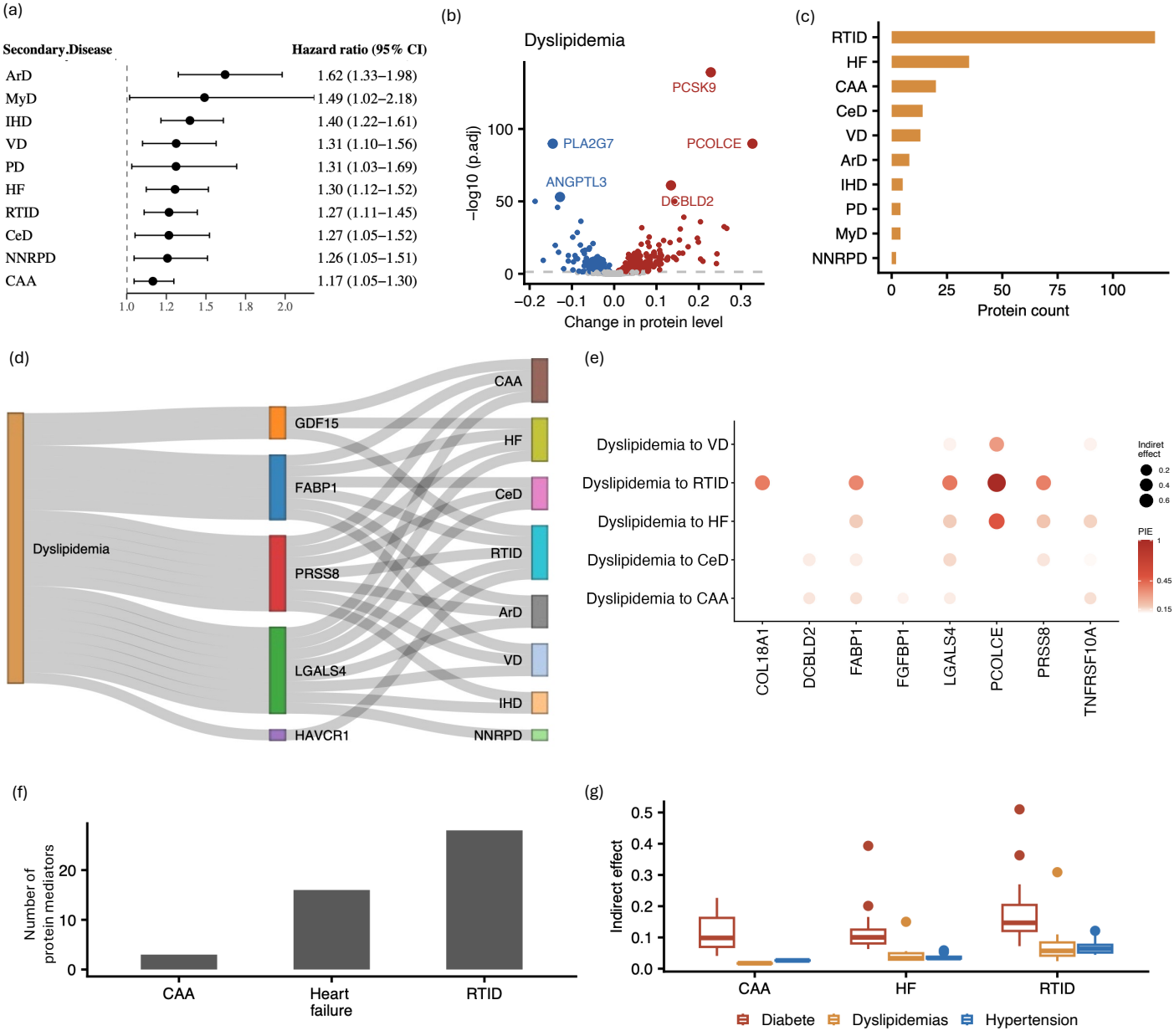
(g)

Stratification analysis

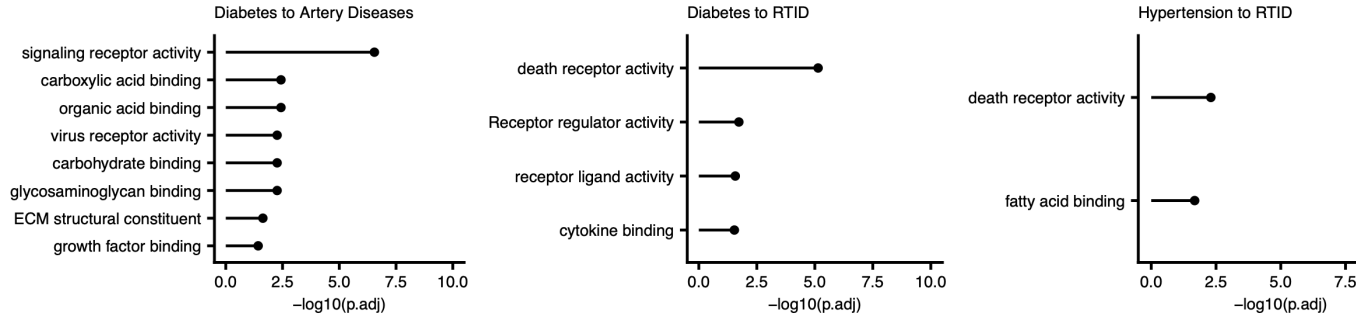



(a)
Secondary.Disease
Hazard ratio (95% CI)
GD 2.50 (1.45–4.31)
RTID 2.03 (1.73–2.40)
CeD 1.86 (1.45–2.38)
HF 1.84 (1.53–2.21)
ArD 1.47 (1.09–1.97)
CAA 1.22 (1.04–1.43)
1.0 2.0 3.0 4.0
(b)
Diabetes
GDF15
ADGRG1
NFASC
PLXNB2
LGALS4
−log10 (p.adj)
Change in protein level
(c)
Diabetes to GD
Diabetes to ArD
Diabetes to CAA
Diabetes to RTID
Diabetes to HF
Diabetes to CeD
Protein count
(d)
Diabetes
FABP1
GDF15
HAVCR1
LGALS4
PRSS8
ArD
CeD
GD
CAA
HF
RTID
(e)
Diabetes to RTID
Diabetes to HF
Diabetes to GD
Diabetes to CeD
Diabetes to CAA
Diabetes to ArD
ADA2
ADAMTSL2
ADGRG1
ANGPT2
GDF15
HAVCR1
LGALS4
PCOLCE
PIGR
PRSS8
REG4
TNFRSF11A
Indiret effect
PIE
(f)
Diabetes.grade
Secondary.disease
Hazard ratio (95% CI)
Severe GD 3.27 (2.24-4.78)
Severe RTID 2.82 (2.52-3.15)
Severe OPDRI 2.24 (1.41-3.56)
Severe CeD 2.13 (1.80-2.53)
Severe HF 1.92 (1.68-2.20)
Severe ArD 1.90 (1.56-2.32)
Severe AD 1.73 (1.32-2.27)
Severe IHD 1.61 (1.40-1.85)
Prediabetes OPDRI 1.58 (1.12-2.25)
Severe VD 1.58 (1.33-1.88)
Severe MyD 1.52 (1.01-2.30)
Prediabetes GD 1.48 (1.05-2.08)
Severe NNRPD 1.44 (1.20-1.73)
Prediabetes RTID 1.41 (1.29-1.54)
Prediabetes ArD 1.35 (1.17-1.55)
Prediabetes MyD 1.34 (1.02-1.76)
Prediabetes VD 1.29 (1.14-1.45)
Severe CAA 1.28 (1.15-1.43)
Prediabetes HF 1.26 (1.14-1.40)
Prediabetes CeD 1.24 (1.09-1.40)
Prediabetes IHD 1.12 (1.02-1.24)
Prediabetes CAA 1.10 (1.02-1.18)
(g)
TNFRSF12A
SERPIND1
REN
REG4
PVR
PRSS8
PRAP1
PLXNB2
PIGR
PCOLCE
OCLN
NFASC
NEFL
MMP12
MFGE8
LGALS4
IL1RL1
HGF
HAVCR1
GFRA1
GDF15
FGF23
FAM3C
FABP1
EDN1
COL6A3
CFI
CFH
C7
BCAN
APOE
ANGPT2
ADGRG1
ADAMTSL2
ADA2
ACE2
Prediabetes to AD
Prediabetes to CAA
Prediabetes to CeD
Prediabetes to GD
Prediabetes to HF
Prediabetes to IHD
Prediabetes to MyD
Prediabetes to OPDRI
Prediabetes to RTID
Prediabetes to VD
Severe to AD
Severe to AlzD
Severe to CAA
Severe to CeD
Severe to GD
Severe to HF
Severe to IHD
Severe to MyD
Severe to NNRPD
Severe to OPDRI
Severe to RTID
Severe to VD
Indiret effect
PIE

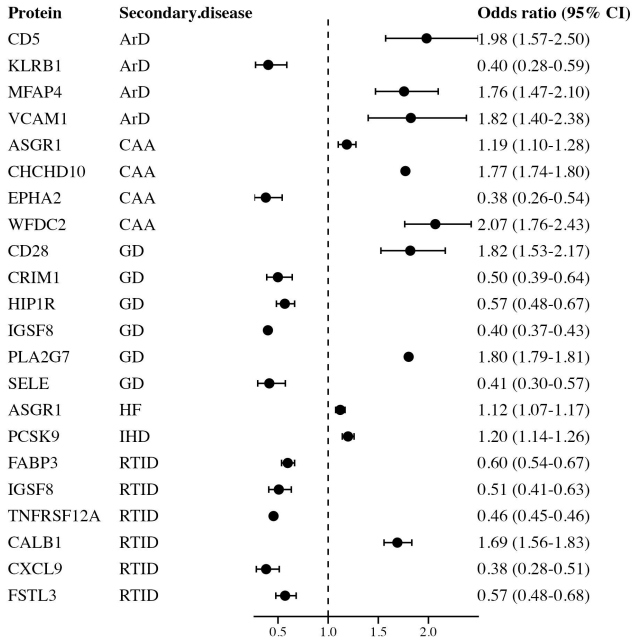

(a)
Secondary.Disease
Hazard ratio (95% CI)
GD 2.34 (1.34–4.08)
OPDRI 2.01 (1.21–3.33)
RTID 1.59 (1.38–1.82)
IHD 1.45 (1.26–1.67)
HF 1.37 (1.16–1.62)
CeD 1.36 (1.12–1.66)
VD 1.34 (1.10–1.62)
CAA 1.23 (1.10–1.38)
1.0 2.0 3.0 4.0
(b)
Hypertension
HAVCR1
ACE2
RBP5
GGT1
LGALS4
−log10 (p.adj)
Change in protein level
(c)
RTID
GD
HF
OPDRI
VD
CeD
CAA
IHD
Protein count
(d)
Hypertension
PRSS8
FABP1
LGALS4
ADM
GDF15
HAVCR1
OPDRI
GD
RTID
HF
CAA
VD
CeD
IHD
(e)
Hypertension to VD
Hypertension to RTID
Hypertension to OPDRI
Hypertension to HF
Hypertension to GD
ADAMTSL2
ADM
ANGPT2
COL18A1
COL6A3
FSTL3
GDF15
GFRA1
HAVCR1
IGFBP4
MMP7
NTproBNP
PALM2
SPON2
PIE
Indiret effect
(f)
Hypertension.grade
Secondary.disease
Hazard ratio (95% CI)
Grade 3 OPDRI 2.37 (1.34-4.19)
Grade 3 CeD 2.14 (1.70-2.69)
Grade 3 IHD 1.59 (1.31-1.93)
Grade 2 CeD 1.58 (1.37-1.82)
Grade 3 VD 1.58 (1.25-1.99)
Grade 3 HF 1.49 (1.22-1.83)
Grade 3 RTID 1.45 (1.22-1.73)
Grade 3 ArD 1.40 (1.05-1.87)
Grade 2 IHD 1.38 (1.23-1.54)
Grade 3 CAA 1.34 (1.16-1.54)
Grade 2 VD 1.29 (1.13-1.49)
Grade 1 CeD 1.28 (1.15-1.44)
Grade 1 IHD 1.24 (1.13-1.34)
Grade 2 HF 1.21 (1.07-1.36)
1.0 2.0 3.0 4.0
(g)
HAVCR1
TNFRSF11B
PALM2
NTproBNP
MMP7
CHI3L1
BST2
# of linked secondary diseases from grade 3 hypertension
(h)
PTPRB
PALM2
NTproBNP
NPPB
MMP7
LTBP2
IGFBPL1
IGF2R
HEG1
HAVCR1
CRIP2
CPM
CD99L2
ACTA2
Grade 3 to CAA
Grade 3 to HF
Grade 3 to IHD
Grade 3 to OPDRI
Grade 3 to RTID
PIE
Indiret effect


(a)
Secondary.Disease
Hazard ratio (95% CI)
ArD 1.62 (1.33–1.98)
MyD 1.49 (1.02–2.18)
IHD 1.40 (1.22–1.61)
VD 1.31 (1.10–1.56)
PD 1.31 (1.03–1.69)
HF 1.30 (1.12–1.52)
RTID 1.27 (1.11–1.45)
CeD 1.27 (1.05–1.52)
NNRPD 1.26 (1.05–1.51)
CAA 1.17 (1.05–1.30)
1.0 1.2 1.5 1.8 2.0
(b)
Dyslipidemia
PCSK9
PLA2G7
PCOLCE
ANGPTL3
DCBLD2
−log10 (p.adj)
Change in protein level
(c)
RTID
HF
CAA
CeD
VD
ArD
IHD
PD
MyD
NNRPD
Protein count
(d)
Dyslipidemia
GDF15
FABP1
PRSS8
LGALS4
HAVCR1
CAA
HF
CeD
RTID
ArD
VD
IHD
NNRPD
(e)
Dyslipidemia to VD
Dyslipidemia to RTID
Dyslipidemia to HF
Dyslipidemia to CeD
Dyslipidemia to CAA
COL18A1
DCBLD2
FABP1
FGFBP1
LGALS4
PCOLCE
PRSS8
TNFRSF10A
Indiret effect
PIE
(f)
Number of protein mediators
CAA
Heart failure
RTID
(g)
Indirect effect
CAA
HF
RTID
Diabete
Dyslipidemias
Hypertension

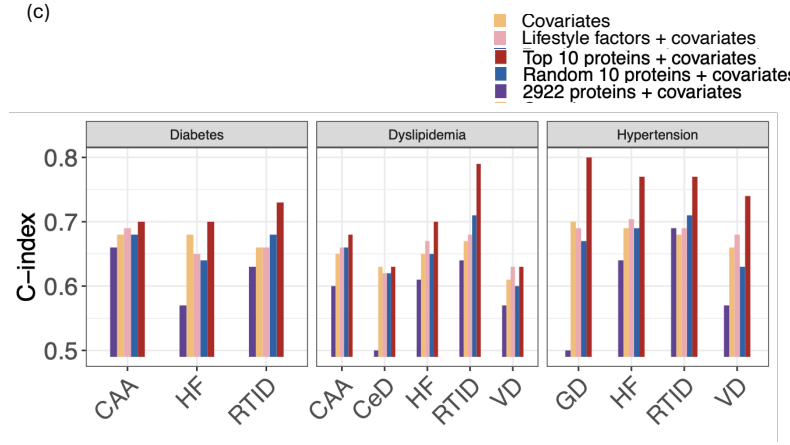
(a)


Diabetes to Artery Diseases
signaling receptor activity
carboxylic acid binding
organic acid binding
virus receptor activity
carbohydrate binding
glycosaminoglycan binding
ECM structural constituent
growth factor binding
0.0 2.5 5.0 7.5 10.0
−log10(p.adj)
Diabetes to RTID
death receptor activity
Receptor regulator activity
receptor ligand activity
cytokine binding
0.0 2.5 5.0 7.5 10.0
−log10(p.adj)
Hypertension to RTID
death receptor activity
fatty acid binding
0.0 2.5 5.0 7.5
−log10(p.adj)

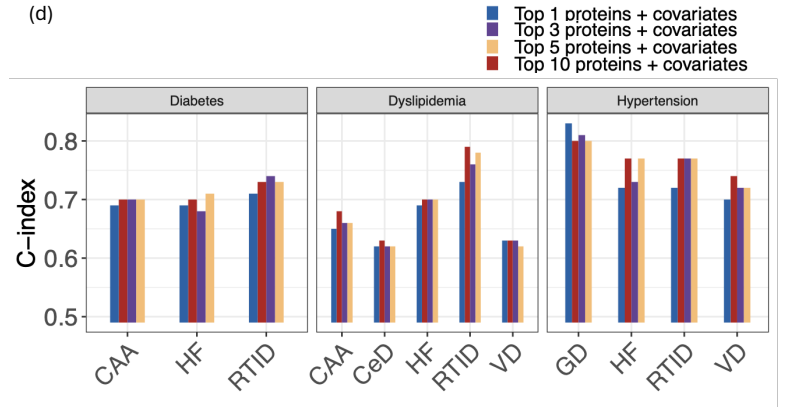
(b)


Protein
Secondary.disease
Odds ratio (95% CI)
CD5 ArD 1.98 (1.57-2.50)
KLRB1 ArD 0.40 (0.28-0.59)
MFAP4 ArD 1.76 (1.47-2.10)
VCAM1 ArD 1.82 (1.40-2.38)
ASGR1 CAA 1.19 (1.10-1.28)
CHCHD10 CAA 1.77 (1.74-1.80)
EPHA2 CAA 0.38 (0.26-0.54)
WFDC2 CAA 2.07 (1.76-2.43)
CD28 GD 1.82 (1.53-2.17)
CRIM1 GD 0.50 (0.39-0.64)
HIP1R GD 0.57 (0.48-0.67)
IGSF8 GD 0.40 (0.37-0.43)
PLA2G7 GD 1.80 (1.79-1.81)
SELE GD 0.41 (0.30-0.57)
ASGR1 HF 1.12 (1.07-1.17)
PCSK9 IHD 1.20 (1.14-1.26)
FABP3 RTID 0.60 (0.54-0.67)
IGSF8 RTID 0.51 (0.41-0.63)
TNFRSF12A RTID 0.46 (0.45-0.46)
CALB1 RTID 1.69 (1.56-1.83)
CXCL9 RTID 0.38 (0.28-0.51)
FSTL3 RTID 0.57 (0.48-0.68)
0.5 1.0 1.5 2.0

(c)


Covariates
Lifestyle factors + covariates
Top 10 proteins + covariates
Random 10 proteins + covariates
2922 proteins + covariates
Diabetes
Dyslipidemia
Hypertension
C−index
0.8
0.7
0.6
0.5
CAA HF RTID
CAA CeD HF RTID VD
GD HF RTID VD

(d)


Top 1 proteins + covariates
Top 3 proteins + covariates
Top 5 proteins + covariates
Top 10 proteins + covariates
Diabetes
Dyslipidemia
Hypertension
C−index
0.8
0.7
0.6
0.5
CAA HF RTID
CAA CeD HF RTID VD
GD HF RTID VD