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 (CVD), cerebral, and renal diseases (RD). However, the underlying molecular mechanisms, particularly the biological paths linking these primary conditions to downstream diseases remain incompletely understood. In this study, we investigated the role of plasma proteins as mediators of secondary disease development using data from ~50,000 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 included GDF15 consistently mediating the diabetes–CVD link and ADM mediating the hypertension–pulmonary disease pathway. Protein mediators of secondary disease development were highly enriched in immune, metabolic, and cytokine-related pathways. Mendelian randomization supported causal roles for 84 proteins, highlighting their potential as therapeutic targets. Moreover, the identified mediating proteins improved predictive accuracy for secondary disease risk compared to other proteins and traditional clinical risk factors from machine learning methods. For example, in individuals with hypertension, the inclusion of top mediating proteins improved prediction accuracy for glomerular disease risk by approximately 14% measured by C-index. Stratified analyses based on disease severity revealed additional disease progression pathways and mediating proteins, such as APOE mediating the association between severe diabetes and Alzheimer's disease. Together, these findings implicate plasma proteins as central molecular mediators linking these primary diseases to subsequent development of a secondary condition 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 chronic conditions, hereafter referred to as primary diseases, not only exert substantial economic and social burdens but also serve as major antecedents to a wide spectrum of downstream health complications. For example, diabetes and hypertension are well-recognized drivers of multi-organ disease progression, contributing to elevated risks of cardiovascular outcomes—including heart failure (HF) and ischemic heart disease (IHD) [2–5] —as well as renal dysfunction and chronic kidney disease [6]. Uncovering the molecular and physiological mechanisms that link these primary diseases to secondary disease development is crucial for improving risk stratification, informing preventive strategies, and guiding the development of targeted interventions, particularly in populations already burdened by these chronic conditions.

Emerging evidence has delineated a complex network of biological mediators that link 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. Among the key pro-inflammatory signals, interleukin-6 (IL-6) and monocyte chemoattractant protein-1 (MCP-1) are consistently upregulated in metabolic disease states and contribute to endothelial activation, monocyte recruitment, and extracellular matrix remodeling, thereby promoting vascular structural changes[7]. Dysregulation of the renin-angiotensin-aldosterone system (RAAS), particularly increased angiotensin II, enhances oxidative stress via NADPH oxidase, activates NF-κB in vascular and immune cells, and upregulates inflammatory cytokines (e.g., TNF-α), chemokines (e.g., MCP-1), and adhesion molecules (e.g., VCAM-1, P-selectin), thereby contributing to hypertension, vascular inflammation, and accelerated atherosclerosis, especially in diabetic and dyslipidemic contexts[8,9]. 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 complete landscape of clinically accessible mediators detectable in blood remains largely unexplored. Clarifying this spectrum is critical, as it holds the potential to transform risk stratification, refine preventive strategies, and enable the development of targeted interventions, particularly for populations facing already high burdens of disease.

Emerging high-throughput plasma proteomic platforms, including Olink[10] and SomaScan[11], now offer the ability to profile 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 [12,13], predicting incident disease and mortality[14–16], and linking diverse phenotypic traits to underlying disease status[17,18]. Here, we harness advanced plasma proteomic technologies within a rigorous causal inference framework[19] to systematically investigate mediating protein pathways underlying the progression from diabetes, hypertension, and dyslipidemia to their respective secondary diseases. This approach aims to bridge molecular discovery with potential clinical application, advancing disease understanding and facilitating the identification of actionable biomarkers for improved patient care.

In this study, we leveraged data from the UK Biobank (UKB), a large, prospective, population-based cohort initiated in 2006 that has enrolled over 500,000 individuals aged 40-69 years at baseline, with ongoing longitudinal follow-up of clinical outcomes[20]. Proteomic measurements were obtained from a randomly selected subset of participants (n = 53,030)[21]. Using these data, we aimed to explore the mediating roles of plasma proteins in the disease paths linking three prevalent conditions, diabetes, hypertension, and dyslipidemia, to 18 secondary disease outcomes categorized into four major clinical domains. For proteins demonstrating evidence of mediation, we further characterized their biological functions, assessed their potential causal effects on secondary diseases, and evaluated their utility in predicting secondary disease risk among individuals with primary diseases. Finally, we stratified participants by the severity of primary diseases using continuous clinical biomarkers to identify additional disease paths and protein mediators, particularly among individuals at higher risk (Fig. 1). By mapping the connections between primary diseases, plasma protein profiles, and secondary disease risk, our study aims to uncover the molecular mechanisms that underlie disease progression and to

identify potential biomarkers that could inform early intervention and therapeutic strategies. Additionally, the mediation-based causal inference framework developed in this study is readily extendable to other primary disease risk factors as well as to other classes of molecular mediators.

Results

Population characteristics and diseases

We included 53,030 participants with an average age of 57.4 years, of whom 53.9% were women (Supplementary Table 1). After data preprocessing (described in the Methods section), expression data for 2,922 plasma proteins and clinical information on three primary diseases, diabetes, hypertension, and dyslipidemia (Supplementary Table 2), as well as 18 secondary diseases (Supplementary Table 3), were retained for analysis to investigate candidate proteins involved in the primary-secondary disease development pathway[21]. In this study, primary diseases were chronic conditions that serve as risk factors for future incident health outcomes, while secondary diseases were potential outcomes resulting from the presence of primary diseases. We examine 18 secondary diseases spanning four major clinical domains, including seven cardiovascular diseases (CVD), two renal diseases (RD), two pulmonary diseases (PuDs), and seven neurological diseases (NDs) (Fig. 1 and Supplementary Table 3). Primary disease status was defined as 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%). Secondary disease incident cases (i.e., number of diagnosed participants after plasma protein collection data) varied from 106 (0.2%) to 8,962 (16.9%) (Supplementary Table 3).

Primary Diseases Increase the Risk of Secondary diseases

All three primary diseases were identified as significant risk factors for a range of secondary diseases, as indicated by hazard ratios (HRs) greater than one in Cox regression models. Dyslipidemia was associated with the widest range, showing significant associations with 10 out of 18 secondary diseases. This was followed by hypertension and diabetes, which were significantly associated with eight and six secondary diseases, respectively (Fig. 2a).

The strongest associations were observed between diabetes and glomerular diseases (GD; HR = 2.501 , $p = 9 \times 10^{-4}$), hypertension and GD (HR = 2.336, $p = 0.002$), and diabetes and

renal tubulo-interstitial disease (RTID; HR = 2.034, $p < 1 \times 10^{-10}$), with both GD and RTID classified as renal diseases, findings that are consistent with prior reports[22–24]. In addition to their strong links with renal outcomes, all three primary diseases showed significant associations with three cardiovascular conditions, cerebrovascular diseases (CeD), cardiac arrest and arrhythmias (CAA), and HF[25,26]. These observations align with existing literature on the systemic impact of these chronic conditions. Notably, we identified disease-specific associations unique to individual primary diseases. For example, nerve, nerve root, and plexus disorders (NNRPD) and Parkinson's disease (PD) were significantly associated with dyslipidemia, but not with diabetes or hypertension.

Primary Diseases are Associated with Plasma Protein Expression

The primary disease occurrence was significantly associated with a wide range of plasma protein expression (with Benjamini-Hochberg adjusted p values < 0.05). Among three primary diseases, diabetes was found to be significantly associated with 1,260 plasma proteins, followed by hypertension (928 proteins), and dyslipidemia (557 proteins) (Fig. 2b). Notably, the majority of protein-disease associations were positive, indicating elevated protein expression levels in individuals with these chronic conditions.

Several strong associations emerged. Diabetes was significantly associated with proteins such as 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, reinforcing the robustness of our findings[27]. Hypertension showed the strongest association with HAVCR1, marked by the lowest p-value ($p < 1 \times 10^{-10}, r = 0.19$, r is the effect size). Gene HAVCR1 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)[28]. For dyslipidemia, PCSK9 exhibited a strong association ($p < 1 \times 10^{-10}, r = 0.23$), consistent with its established role in lipid metabolism and clinical intervention[29]. 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)[29].

Plasma Proteins as Mediators Between Primary and Secondary Diseases

Plasma proteins and secondary diseases that were significantly associated with primary diseases were further examined using mediation analysis to assess whether the effects of primary diseases on secondary outcomes may be mediated, or at least partly mediated through protein-level pathways (Fig. 1c, Methods). To ensure robustness, we applied stringent criteria to identify mediators and retained only mediation pathways that satisfied all of the following: (i) Benjamini–Hochberg adjusted $p < 0.05$ for the indirect effect; (ii) An absolute indirect effect ≥ 0.01; and (iii) A proportional indirect effect (PIE), defined as the ratio of the indirect effect to the total effect, of at least 10%.

We identified a total of 1,461 significant primary disease–plasma protein–secondary disease mediation pathways, involving the three primary diseases, 395 unique plasma proteins, and 13 secondary diseases (Supplementary Table 4). As expected, all indirect effects were positive, with HRs greater than 1, indicating that primary diseases increased the risk of secondary diseases via protein-mediated mechanisms. The majority of mediation effects were observed in the context of diabetes, with 341 distinct proteins mediating associations between diabetes and six secondary diseases: artery diseases (ArD), CeD, GD, HF, CAA, and RTID (Fig. 2c). Dyslipidemia was linked to 131 unique mediating proteins across ten secondary diseases, while hypertension involved 64 unique mediators across seven secondary diseases.

A total of 25 proteins, most notably GDF15, LGALS4, PRSS8, and HAVCR1, consistently mediated all six diabetes-plasma protein-secondary disease mediation pathways (Fig. 2d; Supplementary Fig. 2a), representing the highest number of mediated pathways among all proteins. Many of these high-frequency mediators also exhibited strong and recurrent mediation effects across multiple disease mediation pathways (Fig. 2e; Supplementary Table 5). Among them, GDF15, a well-established biomarker previously linked to diabetes[30,31], emerged as the most prominent mediator, consistently showing the largest indirect effects across all six diabetes-related mediation pathways. In addition, LGALS4 was frequently ranking among the top five mediators in terms of indirect effect size within diabetes-associated mediation pathways.

In the dyslipidemia-related mediation pathways, LGALS4 and PRSS8 exhibited significant mediation effects in eight and seven pathways, respectively, the highest numbers observed among all mediators for dyslipidemia (Fig. 2d, Supplementary Fig. 2b). Notably, LGALS4 also consistently ranked among the top five mediators in multiple dyslipidemia-associated pathways, suggesting a broader functional role across metabolic disease development pathways. In addition to these shared mediators, FABP1, a key protein involved in fatty acid metabolism[32] , emerged as a prominent dyslipidemia-specific mediator, underscoring its potential relevance to lipid-driven disease development pathway (Fig. 2e).

In hypertension-related disease mediation pathways, GDF15 and HAVCR1 emerged as key mediators, mediating six and five secondary disease associations, respectively (Fig. 2d, Supplementary Fig. 2c). In addition to proteins that mediated multiple mediation pathways, we identified several mediators with strong, pathway-specific effects. For instance, Adrenomedullin (ADM) and fatty acid-binding protein 1 (FABP1) represent distinct but biologically plausible mediators linking primary cardiometabolic risk factors to downstream organ disease (Fig. 2e). In hypertension, ADM, which is markedly upregulated as a compensatory vasodilatory and natriuretic peptide that counterbalances excessive vascular tone, attenuates fibrosis, and preserves renal and cardiac function[33], are strongly associated with adverse cardiovascular and renal outcomes. The mediating effect may work through sustain blood pressure elevation rather than directly promoting disease progression. In contrast, 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. Elevated circulating FABP1 has been consistently linked to obesity, insulin resistance, and cardiovascular events, raising the possibility that it operates as both a mechanistic mediator of lipid-driven pathology and a sensitive biomarker of hepatic stress. Together, these pathways highlight how distinct mediator proteins arising from vascular and metabolic origins may contribute to the translation of primary risk factors-hypertension and dyslipidemia-into secondary cardiovascular and renal disease.

We also examined 31 plasma proteins that exhibited significant mediation effects in pathways linking all three primary diseases to CAA, RTID, or heart failure, three secondary diseases whose risks were significantly influenced by all three primary conditions (Fig. 2f;

Supplementary Fig. 3). Notably, GDF15 and IGFBP4 mediated associations between all three primary diseases and each of these three secondary diseases, highlighting their central role in multimorbidity-related pathways. ACE2, a well-established therapeutic target in renal disease [34], specifically mediated associations between all three primary diseases and RTID.

Biological Function Annotation of the Protein Mediators

To understand the enriched biological function of identified plasma proteins mediated in each disease's mediation pathways, we performed pathway enrichment analysis using the GO:BP [35,36] and KEGG [37] databases to annotate the enriched biological processes and pathways[38]. Among the three primary diseases, like the mediation analysis results, plasma proteins mediating diabetes and subsequent diseases are ranked the most, for example, diabetes and GD are enriched in various cytokine and immune response-related pathways. Dyslipidemia and hypertension were also enriched in immune response-related, metabolic, and cytokine-related pathways among proteins mediating multiple disease paths (Fig. 3a), suggesting that primary diseases may influence immune and metabolic functions, thereby increasing susceptibility to secondary conditions. The PI3K-Akt signaling pathway, angiogenesis, lipid transportation biological process and pathways are also enriched for mediators of these three primary disease-initiated disease paths.

Notably, proteins mediating pathways linked to valve disorders (VD) showed significant enrichment in the positive regulation of cardiac contraction pathway, aligning with clinical observations that impaired cardiac regulatory mechanisms contribute to the pathogenesis of valvular disease.

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

We used Mendelian randomization (MR) to investigate whether proteins with significant indirect effects in disease mediation pathways also play causal roles in secondary diseases. MR analysis was conducted for the mediator proteins and linked secondary diseases using genome-wide significant *cis*-protein quantitative trait loci (pQTLs) as instrumental variables[39]. 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. GWAS summary statistics were sourced from the IEU OpenGWAS database[40,41] (See methods).

Among the 341 proteins linking diabetes to six secondary diseases, 69 plasma proteins exhibited evidence of potential causal effects on five outcomes: CereD, GD, CAA, HF, and RTID. 24/131 and 12/64 proteins associated with dyslipidemia and hypertension, respectively, had significant MR evidence of causal roles (Supplementary Table 6). Based on outcomes, CAA was linked to the largest number of putative causal proteins (38) at Bonferroni-adjusted $p < 0.05$. Notably, all these 38 proteins were significant mediators in the diabetes-CAA pathway, and three were also identified as mediators in the dyslipidemia-CAA pathway. No causal mediators were identified for the hypertension-CAA path.

Our analysis confirmed several established causal roles of proteins in disease mediation pathways (Fig. 3b). For example, we found that VCAM1 is potentially causal associated with artery diseases (OR = 1.82, $p = 1.36 \times 10^{-7}$), which aligns VCAM1's function that VCAM1 as a key endothelial adhesion molecule mediating leukocyte recruitment to sites of vascular inflammation, thus contributes to plaque formation, progression, and instability, making it a central player in the development of atherosclerosis and related vascular complications[42]. ASGR1, a key protein involved in cholesterol metabolism, showed significant mediation effects in pathways from diabetes, hypertension, and dyslipidemia to multiple types of CVDs. It also showed potential causal roles in two CVDs, including CAA (OR = 1.186, $p = 7.06 \times 10^{-6}$) and HF (OR = 1.121, $p = 1.01 \times 10^{-7}$).

PCSK9 inhibitors, such as evolocumab and alirocumab (monoclonal antibodies), are known to robustly reduce LDL-C levels and lower the risk of major adverse cardiovascular events (MACE). In our study, PCSK9 was identified as a potentially causal factor for several cardiovascular diseases, including angina pectoris (IVW OR = 1.27, p = 4.44×10^{-13})[43]. Moreover, PCSK9 consistently exhibited significant mediation effects in disease development pathways linking dyslipidemia to three types of CVDs: heart failure (HF), valve disorders (VD), and ischemic heart disease (IHD).

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

Utilizing a machine learning model, we further assessed their predictive power for secondary diseases among participants already diagnosed with a primary disease. This analysis focused on disease mediation pathways involving at least 10 proteins. We found that using the top ten proteins ranked by mediation effect size as predictors, consistently achieved superior predictive accuracy (measured by C-index, see Methods) compared to models based on (i) basic demographic confounders (age, sex, BMI), (ii) lifestyle and clinical biomarkers including blood pressure, cholesterol, smoking and drinking; (iii) randomly selected sets of 10 proteins (repeated sampling, n = 100), and (iv) the full proteomic panel, underscoring the predictive relevance of proteins identified via mediation analysis based on machine learning models (Fig. 3c and Supplementary Table 7). Notably, models using either the entire proteomic profile or random subsets of proteins frequently demonstrated lower predictive accuracy across multiple secondary diseases. This highlights the detrimental impact of including non-informative or weakly relevant proteins and emphasizes the importance of biologically informed feature selection in improving model performance. These findings support the clinical value of mediator proteins as potential targeted biomarkers for forecasting future disease risk in individuals with diabetes, hypertension, or dyslipidemia.

We further investigated how the number of proteins included in the models influenced prediction accuracy. In 10 out of 12 tested diseases mediation pathways, the LASSO model using the top 10 plasma proteins outperformed models using only the top five, three, or a single protein. Nonetheless, in most of these cases, the predictive performance of models using the top 5 proteins was comparable to that achieved using the top 10, indicating that a smaller, more parsimonious set of high-impact proteins may still yield clinically relevant prediction performance while improving interpretability and model efficiency.

Blood Pressure–Based Stratification Reveals Similar Disease Development Pathways and Mediating Proteins as Diagnostic History

Hypertension control is an important determinant of health outcomes, as sustained elevations in blood pressure are associated with increased risk of adverse outcomes. To assess

whether participants at different levels of blood pressure have different risks for secondary diseases, as well as distinct patterns of protein mediators along these disease pathways, we further stratified participants by their 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). This stratification approach was based on the 2023 European Society of Hypertension (ESH) guidelines [44] and is described in detail in the Methods section.

We identified eight significant primary disease–secondary disease development pathways. In each case, participants in at least one hypertension category, always including grade 3, exhibited a significantly higher risk of developing secondary diseases compared to those with normal blood pressure (Fig. 4a). Seven of these eight secondary diseases overlapped with the disease list identified from survival analysis using hypertension diagnosis history as exposure. The only newly identified disease development pathway was a significantly increased risk of ArD among participants with grade 3 hypertension (HR = 1.401, 95% CI: 1.052–1.866), a path not observed in participants with grades 1 or 2 hypertension or in those classified by diagnosis history. As expected, secondary diseases associated with lower levels of hypertension were consistently retained among the paths identified at higher levels, with the magnitude of risk generally increasing with hypertension severity for the overlapping secondary diseases. The risk profile of participants with diagnosed hypertension generally aligned between those in grades 2 and 3.

A total of 197 plasma proteins were identified as mediators in the relationship between at least one grade of hypertension severity and a secondary disease (Supplementary Table 8 and Supplementary Table 9). Participants with grade 3 hypertension were more likely to exhibit protein-mediated increases in secondary disease risk compared to those with lower hypertension levels, as evidenced by both a broader range of mediating proteins (162 unique proteins, Fig. 4b) and stronger indirect effects through overlapping plasma proteins linked to the same secondary diseases (Fig. 4c). Many of the key mediators linked to grade 3 hypertension overlapped with those identified when hypertension status was defined based on diagnostic history (Fig. 4d, e and Supplementary Table 9). For example, HAVCR1 was associated with the largest number of secondary diseases (n = 6) and also exhibited strong mediation effects, while MMP7

demonstrated significant indirect effects across five disease mediation pathways under both classification strategies. Additionally, NT-proBNP exhibited particularly strong mediation effects for several cardiovascular outcomes, including HF (HR = 1.16, PIE = 38%) [45].

Nevertheless, we also identified proteins uniquely associated with grade 3 hypertension that were not detected when using diagnostic history alone. For example, CRIP2, a protein highly expressed during cardiovascular development, emerged as a significant mediator in the grade 3 hypertension analysis but not under the diagnosis-based definition. Furthermore, comparative analyses of overlapping proteins linked to the same secondary diseases revealed that proteins identified through grade 3 hypertension consistently exhibited stronger indirect effects than those identified via hypertension diagnosis (Fig. 4f). Notably, several of these newly identified mediators also showed evidence of potential causal effects on secondary diseases, as supported by Mendelian randomization analysis.

Glycated Hemoglobin–Based Stratification Uncovers Novel Disease Development Pathways Beyond Diabetes History

Similarly, we classified all participants into three categories based on glycated hemoglobin (HbA1c)[46] to define diabetes status: normal (HbA1c < 5.7%), prediabetes (5.7% ≤ HbA1c < 6.4%), and severe (HbA1c ≥ 6.4%). This approach revealed six new disease development pathways that were not identified when using diagnosis history alone (Fig. 5a). Newly identified paths included three CVDs, IHD, myocardial disorder (MyD), and valve disorder; two NDs (Alzheimer's disease, AD, and NNRPD); and one PD (OPDRI). These findings suggest that while a diabetes diagnosis is indicative of health risk, failure to control HbA1c within normal ranges may confer an even greater vulnerability to developing a broader and potentially more severe set of secondary diseases. Additionally, participants with sever diabetes exhibited significantly higher risks for AD[47], which aligns with clinical observations.

Unlike the analysis of stratified hypertension severity, where few proteins mediated the pathways between lower-grade hypertension and secondary diseases, we observed a substantial number of mediating proteins associated with prediabetes (Supplementary Table 10 and Supplementary Table 11). Specifically, 394 proteins mediated the relationship between

prediabetes and 10 secondary diseases, exceeding the number identified for severe diabetes (352 proteins linked to 12 secondary diseases) and for diabetes defined by diagnosis history (341 proteins), with 199 proteins shared across these classifications (Fig. 5b, and Supplementary Fig. 4a). When comparing the magnitude of indirect effects for overlapping proteins linked to the same secondary diseases, severe diabetes consistently exhibited stronger mediation effects, followed by diabetes diagnosis history (Fig. 5c).

Although most of the important protein mediators (identified using the previously described criteria) overlapped with those identified based on diagnosed history (Supplementary Fig. 4b, c and Supplementary Table 11), we also observed notable differences in mediation patterns. For example, GDF15 and LGALS4 exhibited significant indirect effects in 21 disease mediation pathways, the highest among all plasma proteins, mirroring results obtained using diabetes diagnostic status. However, GDF15 showed strong mediation effects only in severe diabetes and diabetes diagnosis–based classifications, but not in prediabetes (Fig. 5d). In contrast, LGALS4 consistently acted as a strong mediator across all three scenarios. In addition, several new protein mediators were identified in disease paths unique to stratified diabetes levels. For instance, APOE, PVR, and NEFL were implicated in the pathway to AD, with APOE [48] being a well-established biomarker for neurological disorders. Similarly, CFI, CFH, and SERPIND1 emerged as strong mediators in pathways leading to IHD, highlighting additional molecular links between glycemic dysregulation and cardiovascular risk.

As we found the importance of blood glucose management in reducing sequent disease incidence, we try to further evaluate an optimal HbA1c control level to preventing sequent disease. We used a generalized additive model (GAM) to assess the association between HbA1c levels and the risk of multiple secondary diseases[49]. For AD, we observed a sharp increase in risk when HbA1c levels exceeded approximately 5.7-5.8%, with a steeper rise beyond 7.9% (Fig. 5e), which align with the clinical practice and enhance the importance of maintain normal blood glucose (HbA1c < 5.7%), the sequent disease risk dramatical increase even in the moderate blood glucose range. In contrast, risk remained relatively low at HbA1c levels below 5.7%. Additionally, uncertainty, reflected by wider confidence intervals, increased markedly at extremely high HbA1c values, underscoring the importance of maintaining glycemic control within a normal range to reduce the risk of adverse outcomes. These results suggest the presence

of two critical HbA1c “acceleration points”, around 5.7% and 7.9%, beyond which the risk of AD increases substantially.

A distinct pattern was observed for valve disorders, where the lowest risk occurred near an HbA1c level of 5.2%, followed by a pronounced increase in risk beyond approximately 5.8% (Fig. 5f). Interestingly, a modest elevation in risk was also noted at very low HbA1c levels; however, this may be attributed to statistical variability at the extremes of the HbA1c distribution, suggesting that both hyperglycemia and hypoglycemia warrant further investigation. For most secondary diseases, we observed a largely linear relationship between disease risk and HbA1c levels within the intermediate range, reinforcing the clinical relevance of HbA1c as a continuous biomarker for disease risk stratification.

Discussion

In summary, leveraging data from 53,030 UK Biobank participants profiled with plasma proteomics and linked clinical records, we systematically identified disease development pathways linking three common risk factors, diabetes, hypertension, and dyslipidemia, to 18 secondary diseases spanning five major organ systems. Building on these identified paths, we investigated the potential mediating role of plasma proteins in the progression from primary to secondary diseases. Biological pathway enrichment analysis revealed significant enrichment of immune response-related, metabolic, and cytokine related pathways among proteins mediating multiple disease paths. Many of the mediating proteins were further supported as causal factors through MR analysis. Moreover, the identified mediating proteins offered stronger predictive accuracy for secondary disease risk compared to other proteins and traditional clinical confounders from machine learning methods. For example, in individuals with hypertension, the inclusion of top mediating proteins improved prediction accuracy (measured by C1-index) for glomerular disease risk by approximately 14%. Finally, stratifying participants by disease severity using relevant clinical biomarkers, blood pressure, and HbA1c, revealed additional disease development pathways and protein mediators. Together, our findings provide a comprehensive map of protein mediators across disease mediation pathways, highlighting a subset of proteins with potential as therapeutic targets and biomarkers.

Our findings support and extend previous observations linking the three primary diseases, diabetes, hypertension, and dyslipidemia, to a broad range of secondary diseases. Numerous

studies have established these conditions as key risk factors CVDs and RDs[5,22,24–26], with dyslipidemia also implicated in PD[50]. Our results confirmed these well-known associations and further uncovered several unexpected relationships, such as a novel link between hypertension and a subtype of PuD (OPDRI). In addition, we validated many established associations between diseases and specific plasma proteins, including the link between diabetes and GDF15, and between dyslipidemia and PCSK9.

Our results suggest that plasma proteins play a mediating role in the progression from primary to secondary diseases. Specifically, we identified 395 unique proteins that mediated associations involving 13 secondary diseases in 1,461 significant primary mediation pathways. These mediators included proteins that consistently exhibited strong effects linking a single primary disease to multiple secondary outcomes, for example, GDF15 and LGALS4, both of which strongly mediated the relationship between diabetes and six different secondary diseases. We also identified proteins that mediated associations from all three primary diseases to a single secondary disease, such as ASGR1 in the mediation of heart failure. MR analysis further supported the potential causal roles of many of these "strong" mediators in influencing secondary disease risk. Together, these findings provide a list of candidate proteins with potential therapeutic relevance for individuals with primary diseases, offering opportunities for targeted intervention to prevent downstream disease development.

Complementing these findings, machine learning models demonstrated that the identified strong mediating proteins provided superior predictive power for secondary disease risk among individuals with primary diseases, outperforming both other plasma proteins and conventional clinical confounders. Notably, some proteins showed both evidence of causal effects and high predictive utility. For example, HAVCR1, which mediated the pathway from hypertension to valve disorder, emerged as a key predictor of secondary disease risk, highlighting its potential relevance for both mechanistic insight and risk stratification.

Beyond defining primary disease status based on clinical diagnoses, we also considered disease severity by incorporating continuous clinical biomarkers, specifically, blood pressure and HbA1c levels, allowing us to stratify participants into multiple risk categories. HbA1c-based stratification revealed several novel disease development pathways not captured through diagnosis-based classification, including a previously unreported association between poor glycemic control and increased risk of AD. In hypertension, stratification by blood pressure

levels yielded a disease development pathways largely consistent with that based on hypertension diagnostic history. By further modeling the nonlinear relationship between HbA1c and disease risk, we identified two potential "acceleration points": 5.8% and 7.9%, beyond which the risk of AD increased sharply, which provides more evidence for diabetes patients' blood glucose management. These stratifying findings demonstrate the importance of blood glucose and pressure management in preventing further clinical outcomes in at-risk populations. We did not stratify dyslipidemia severity based on biomarkers such as lipoprotein(a) because dyslipidemia can be associated with either elevated or reduced lipid levels. In particular, analysis of the UK Biobank data revealed no significant differences in lipoprotein levels between participants with and without a dyslipidemia diagnosis (two-sample t-test), limiting the utility of lipid biomarkers as stratification variables in this context.

Importantly, the mediation-based causal inference framework developed in this study is readily extendable to other primary disease risk factors—such as liver disease, a known contributor to CVDs—as well as to other classes of molecular mediators, including plasma metabolites. Building on the identified protein mediators, our framework incorporates downstream statistical analyses for functional annotation, Mendelian randomization to prioritize potential causal targets, and machine learning models to enhance clinical risk prediction. Future research should leverage and expand this integrative framework to further elucidate mechanistic pathways and identify therapeutic opportunities across a broad range of disease contexts.

Several limitations of our study warrant consideration. First, although we examined a broad range of secondary diseases, some conditions had relatively few cases, down to as low as 0.2% prevalence, potentially limiting the reliability of conclusions drawn for these low-prevalence outcomes. Second, the mediation effects of plasma proteins were analyzed individually, without accounting for potential joint or interactive effects among proteins, which may underestimate the complexity of the underlying molecular mediation pathways. Additionally, we assumed a linear relationship for the indirect effects of proteins on secondary disease risk, which may oversimplify the true, potentially non-linear biological processes. 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, highlighting the need for validation in more diverse populations.

In conclusion, leveraging data from approximately 50,000 individuals in the UK Biobank cohort, we investigated the role of plasma proteins in mediating the progression from primary to secondary diseases. A subset of these proteins demonstrated potential as candidate drug targets and clinical biomarkers. These findings prioritize future research aimed at elucidating the molecular mechanisms through which these proteins influence disease progression and at developing targeted therapies to modulate protein levels and prevent downstream diseases in individuals with primary diseases.

Methods

UK Biobank Plasma Protein Cohorts

Proteomic profiling of blood samples from 54,219 participants in the UK Biobank was conducted using the antibody-based Olink Explore 3072 Proximity Extension Assay (PEA), which quantified 2,923 unique plasma proteins (Fig. 1 and Supplementary Table 1). This cohort comprised 46,595 (85.9%) randomly selected participants, 6,376 (11.8%) individuals selected by a consortium, and 1,268 (2.3%) participants from a COVID-19 repeat imaging study [21,51]. To mitigate any potential impact of COVID-19 on plasma protein levels, data from only the first two groups, sampled between March 2006 and September 2010, were utilized. The time plasma proteins data collected was defined as instance 0.

Additionally, a subset of participants was recruited from 2015 to 2019 and again in 2021 for two further rounds of plasma protein measurement. However, due to the limited sample size in these additional measurements (N = 1,173, with only 506 overlapping with the primary analysis cohort), this data was not included in subsequent analyses. Furthermore, the plasma protein GLIPR1 was removed from the analysis due to insufficient data availability.

Primary Diseases and Secondary diseases

The primary diseases analyzed in this study included diabetes, hypertension, and dyslipidemia (Supplementary Table 2), all of which are known to be significant causal factors for multiple subsequent diseases [52–55]. We examined 18 secondary diseases in this analysis (see Fig. 1 and Supplementary Table 3). Due to the low prevalence of some rare diseases within the cohorts, each subsequent disease category encompassed several subtypes. The categorization

method adhered to the original UK Biobank disease categorization table and was informed by professional medical input. Detailed categorization, corresponding ICD-10 codes, and disease prevalence are outlined in Supplementary Table 3.

Primary disease status was defined as binary outcomes, indicating whether participants were diagnosed with a primary disease within the period spanning two years before to one year after the plasma protein data collection. Individuals diagnosed with a primary disease more than two years before instance 0 were excluded to ensure that primary disease status closely reflected the time of plasma protein measurement. The number of positive cases per 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 a primary disease were excluded, as our analysis focused on the future risk of secondary diseases. Follow-up time and outcome data were obtained from the UK Biobank. The follow-up period for each incident outcome began at instance 0 and continued until the earliest occurrence of the outcome or the study's end date, 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 nine covariates are recognized as significant confounders for certain diseases, such as cardiovascular disease. Smoking and alcohol consumption data were gathered through questionnaires administered at the baseline (instance 0), which categorized participants' smoking status into three levels (never, past, and current) and alcohol consumption status into five levels (never, monthly or less, 2 to 4 times a month, 2 to 3 times a week, and 4 or more times a week). Blood pressure (SBP and DBP) and glycated hemoglobin levels were also measured at baseline. To avoid over-adjustment, we excluded covariates that directly overlapped with the definition of primary diseases. Specifically, SBP and DBP were omitted in models where hypertension was the primary disease, as the inclusion could underestimate effect of hypertension on secondary outcomes. HbA1c remained included in these models to account for potential confounding by diabetes. Conversely, when diabetes was the primary disease, HbA1c was excluded, while SBP and DBP were retained.

Mediation Analysis

All statistical analyses were conducted using R version 4.4.2. The same analysis pipeline was applied for each primary-secondary disease development pathway. Survival models were used to estimate the risk of progression from primary diseases to subsequent diseases, adjusting for relevant confounders as specified in Equation (1):

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

Here $H(t)$ is the hazard function at time t, c is the log relative hazard (defined as total effect), X is the binary primary disease status (diagnosed vs. not diagnosed), Ω is the vector of covariates, z_1 are the coefficients for covariates, and T is the transpose of vector. Secondary diseases that demonstrated a statistically significant association with primary diseases ($p < 0.05$) were advanced to mediation analysis.

The relationship between plasma protein levels and the occurrence of primary disease was assessed using linear regression, adjusted for confounders consistent with those in the survival models:

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

Here, M is the expression level of protein i, a is the estimated change in protein level associated with primary disease status, and z_2 represents the coefficients for covariates. Plasma proteins were considered to be significantly associated with the primary disease if the adjusted p-values, corrected using the Benjamini-Hochberg (BH) method, were lower than 0.05.

Mediation analysis was conducted to examine the potential mediating role of plasma proteins in the link between primary and secondary disease occurrences. The mediation analysis using survival models was based on the secondary diseases and plasma proteins significantly associated with primary diseases. In the framework of mediation analysis, the onset of primary disease was considered the exposure variable, the concentration of 2922 plasma proteins measured at baseline being potential mediators respectively, and the occurrence and timing of subsequent diseases as outcome variables[56]. This analysis was also adjusted for previously mentioned confounders:

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

Here, c' represents the direct effect of the primary disease on the secondary outcome, b represents the effect of the mediator (one plasma protein), and $z_3^T\omega$ adjusts for the same confounders.

The indirect effect of each protein was estimated using the product of coefficients method ($a \cdot b$), and its statistical significance was assessed via 1,000 bootstrap resamples[57]. To control for false discovery rates due to multiple testing on 2,922 plasma proteins, the resulting p-values for indirect effects were adjusted using the Benjamini-Hochberg (BH) method[58]. Plasma proteins were classified as mediators between primary and secondary diseases if they exhibited a significant indirect effect, defined as an absolute value of the indirect effect ≥ 0.01 and a proportional indirect effect (PIE), defined as the ratio of the indirect to the total effect, of at least 10%.

We further define the "important protein mediators" by three criteria: (i) plasma proteins that consistently mediated pathways from a single primary disease to multiple secondary diseases; (ii) plasma proteins that exhibited strong mediation effects within an individual primary–secondary disease mediation pathway; and (iii) plasma proteins that mediated pathways linking all three primary diseases to a single secondary disease.

Biological Function Enrichment Analysis

Biological function enrichment analysis was performed (using the Gene Ontology Biological Process, GO:BP, database) on the identified plasma proteins mediators to annotate their biological functions [59]. The pathway enrichment analysis was performed using the GO:BP and KEGG databases [35,36,38]. FDR correction was applied to identify significantly enriched biological processes and pathways. For GO:BP, we additionally identified key driver terms to highlight core functional themes within the enriched gene sets [60].

Mendelian Randomization

We employed conventional two-sample Mendelian Randomization (MR) methods to investigate the causal relationship between identified mediating plasma proteins and corresponding secondary diseases, using genetic variants as instrumental variables [39]. Genetic associations with plasma proteins (pQTLs) were obtained from the UK Biobank European cohort plasma protein dataset, with significance set at $p < 5 \times 10^{-8}$ [61]. Similarly, associations of these

genetic variants with the respective secondary diseases were derived from publicly available IEU GWAS databases, also with $p < 5 \times 10^{-8}$, and restricted to the European cohort [62,63]. The two-sample MR analysis was performed using the R TwoSampleMR package [64,65].

Predicting the Future Risk of Secondary diseases

To examine the predictive power of identified plasma protein mediators, we used plasma proteins with the ten largest mediator effects as inputs in a lasso survival model[66] to predict the future risk of corresponding secondary diseases in participants diagnosed with primary diseases (i.e., one model for each disease pathway). The hyperparameters of the lasso survival model were optimized through cross-validation, and prediction accuracy was assessed using the C-index, calculated from 10-fold cross-validation.

The predictive ability of this lasso model was compared with three alternatives: (1) a lasso survival model that included only age, sex, BMI, and race as inputs; (2) a lasso survival model incorporating clinical and lifestyle biomarkers including age, sex, BMI, blood pressure, HbA1c level, smoking and drinking statues; (3) a lasso survival model using ten randomly selected proteins, with this random selection process repeated 100 times to calculate the average prediction accuracy using the C-index; (4) an XGBoost survival model that used all 2,922 plasma proteins as inputs. XGBoost was chosen for the third model due to its robustness in handling the large proportion of missing values encountered when using all plasma proteins as inputs.

Stratifying Patients by Blood Pressure and Glycated Hemoglobin

In addition to using the history of primary disease diagnoses, we stratified participants into multiple levels based on systolic and diastolic blood pressure (SBP and DBP) as well as glycated hemoglobin (HbA1c) levels. Unlike historical disease data, blood pressure and glycated hemoglobin measurements were taken on the same day as the plasma protein data collection (instance 0). However, these physiological markers are more susceptible to environmental or individual factors, such as medication use and recent physical activity. Thus, in this project, blood pressure and glycated hemoglobin levels were used as supplementary indicators to the history of disease diagnoses.

According to the 2023 European Society of Hypertension (ESH) guidelines, 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) [67]. Similarly, participants were classified into three categories for diabetes based on HbA1c levels: normal (HbA1c < 5.7%), prediabetes (5.7% ≤ HbA1c < 6.4%), and severe (HbA1c ≥ 6.4%)[46].

Based on this categorization, we examined whether participants with hypertension (grades 1 to 3) or diabetes (prediabetes and severe) had a higher likelihood of developing subsequent diseases compared to those with normal levels. Mediation analysis was then performed to assess the potential mediating role of plasma proteins in the associations between varying levels of disease severity and the occurrence of secondary diseases.

Data availability

All UK Biobank data (electronic health record, plasma proteome profile) 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, including those used to replicate the analyses and figures presented in this manuscript.

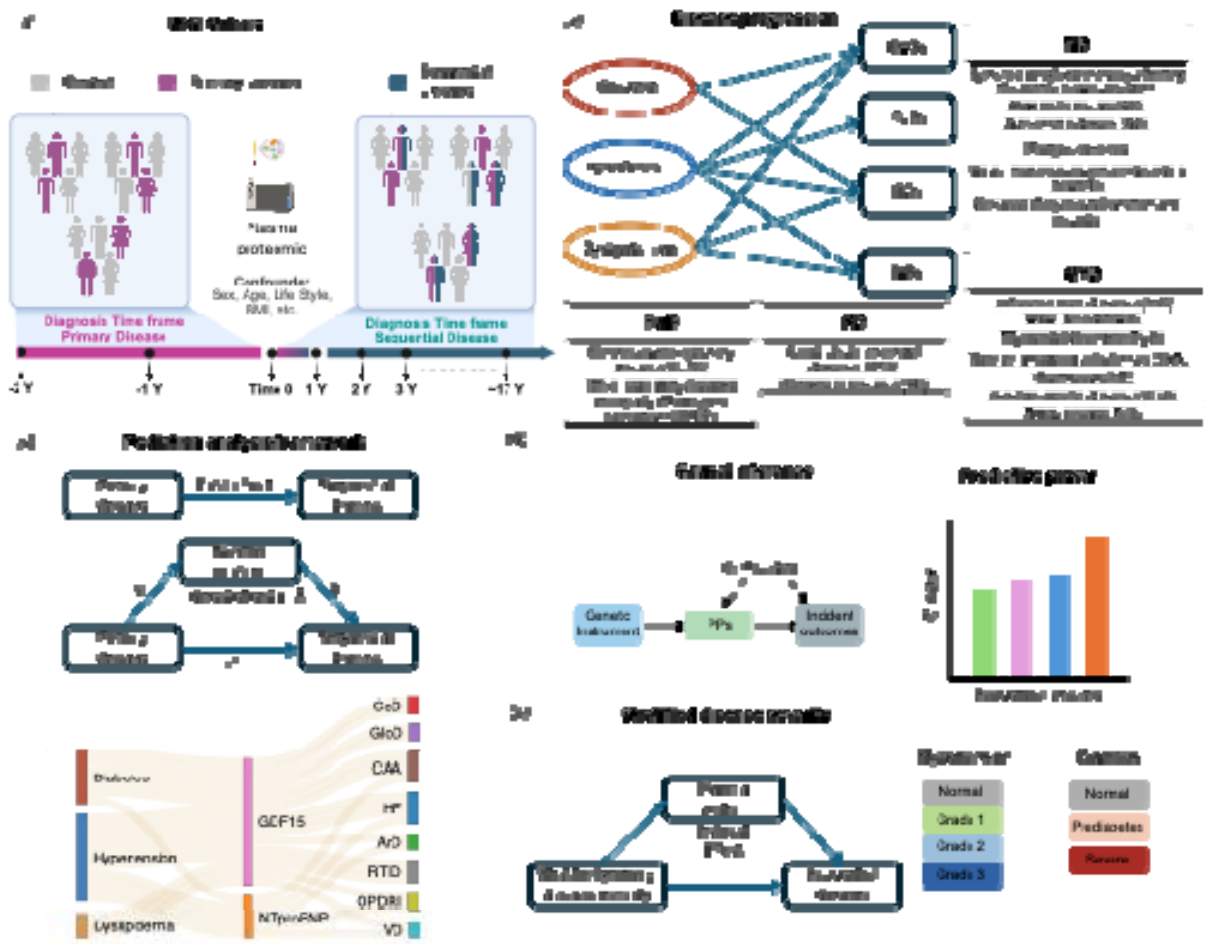
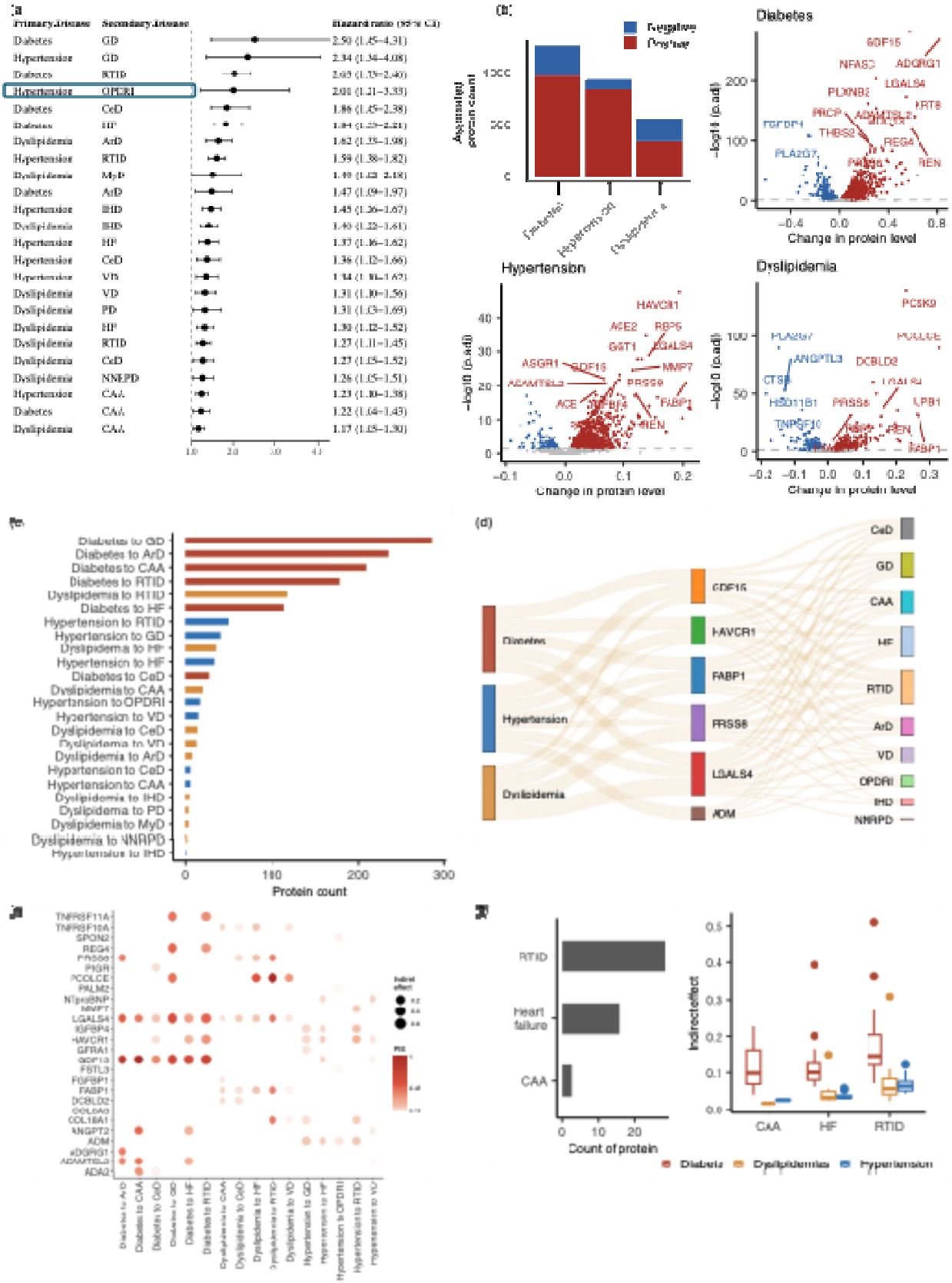


Fig. 1. **Schematic overview of the study design. (a)** The mediation analysis was conducted using data from 53030 UK Biobank participants who completed plasma proteomic profiling. (b) Three primary diseases, diabetes, hypertension, and dyslipidemia, were evaluated as potential risk factors for 18 secondary diseases, grouped into four clinical domains. (c) Mediation analysis was used to identify plasma proteins that mediated the associations between primary and secondary diseases. (d) Identified mediating proteins were further analyzed to annotate their biological functions, assess potential causal effects through genetic analyses, and evaluate their predictive power for secondary disease risk. (e) Primary diseases were additionally stratified by HbA1c and blood pressure levels to examine how disease severity influences secondary disease risk and to identify mediating proteins specific to more severe cases. Figure created with BioRender.com.



(a)
Primary Disease
Secondary Disease
Hazard ratio (95% CI)
Diabetes GD 2.50 (1.45–4.31)
Hypertension GD 2.34 (1.34–4.08)
Diabetes RTID 2.03 (1.73–2.40)
Hypertension OPDRI 2.01 (1.21–3.33)
Diabetes CeD 1.86 (1.45–2.38)
Diabetes HF 1.84 (1.23–2.21)
Dyslipidemia ArD 1.62 (1.33–1.98)
Hypertension RTID 1.59 (1.38–1.82)
Dyslipidemia MyD 1.49 (1.02–2.18)
Diabetes ArD 1.47 (1.09–1.97)
Hypertension IHD 1.45 (1.26–1.67)
Dyslipidemia IHD 1.40 (1.22–1.61)
Hypertension HF 1.37 (1.16–1.62)
Hypertension CeD 1.36 (1.12–1.66)
Hypertension VD 1.34 (1.10–1.62)
Dyslipidemia VD 1.31 (1.10–1.56)
Dyslipidemia PD 1.31 (1.03–1.69)
Dyslipidemia HF 1.30 (1.12–1.52)
Dyslipidemia RTID 1.27 (1.11–1.45)
Dyslipidemia CeD 1.27 (1.03–1.52)
Dyslipidemia NNRPD 1.26 (1.05–1.51)
Hypertension CAA 1.23 (1.10–1.38)
Diabetes CAA 1.22 (1.04–1.43)
Dyslipidemia CAA 1.17 (1.03–1.30)
(b)
Negative
Positive
Diabetes
Hypertension
Dyslipidemia
Diabetes
Hypertension
Dyslipidemia
Change in protein level
−log10 (p.adj)
(c)
Diabetes to GD
Diabetes to ArD
Diabetes to CAA
Diabetes to RTID
Dyslipidemia to RTID
Diabetes to HF
Hypertension to RTID
Hypertension to GD
Dyslipidemia to HF
Hypertension to HF
Diabetes to CeD
Dyslipidemia to CAA
Hypertension to OPDRI
Hypertension to VD
Dyslipidemia to CeD
Dyslipidemia to VD
Dyslipidemia to ArD
Hypertension to CeD
Hypertension to CAA
Dyslipidemia to IHD
Dyslipidemia to PD
Dyslipidemia to MyD
Dyslipidemia to NNRPD
Hypertension to IHD
Protein count
(d)
Diabetes
Hypertension
Dyslipidemia
GDF15
HAVCR1
FABP1
PRSS8
LGALS4
ADM
CeD
GD
CAA
HF
RTID
ArD
VD
OPDRI
IHD
NNRPD
(e)
TNFRSF11A
TNFRSF10A
SPON2
REG4
PRSS8
PIGR
PCOLCE
PALM2
NTproBNP
MMP7
LGALS4
IGFBP4
HAVCR1
GFRA1
GDF15
FSTL3
FGFBP1
FABP1
DCBLD2
COL18A1
ANGPT2
ADM
ADGRG1
ADAMTSL2
ADA2
Indirect effect
PIP
(f)
RTID
Heart failure
CAA
Count of protein
Indirect effect
CAA
HF
RTID
Diabetes
Dyslipidemias
Hypertension

Fig. 2. Mediation analysis results. (a) Hazard ratios and 95% confidence intervals for the associations between three primary diseases and 18 secondary diseases; only statistically significant associations are shown ($p < 0.05$). (b) Associations between the three primary diseases and plasma proteins; most proteins showed positive associations. (c) Number of plasma proteins exhibiting significant mediation effects in each primary–secondary disease path. (d) Plamsa proteins mediated in the disease development pathway from each primary disease to multiple secondary diseases. (e) Dot plot showing the top five plasma proteins with the strongest mediation effects in each disease mediation pathway. (f) Comparison of mediation effects for proteins that mediated associations from all three primary diseases to a single secondary disease.

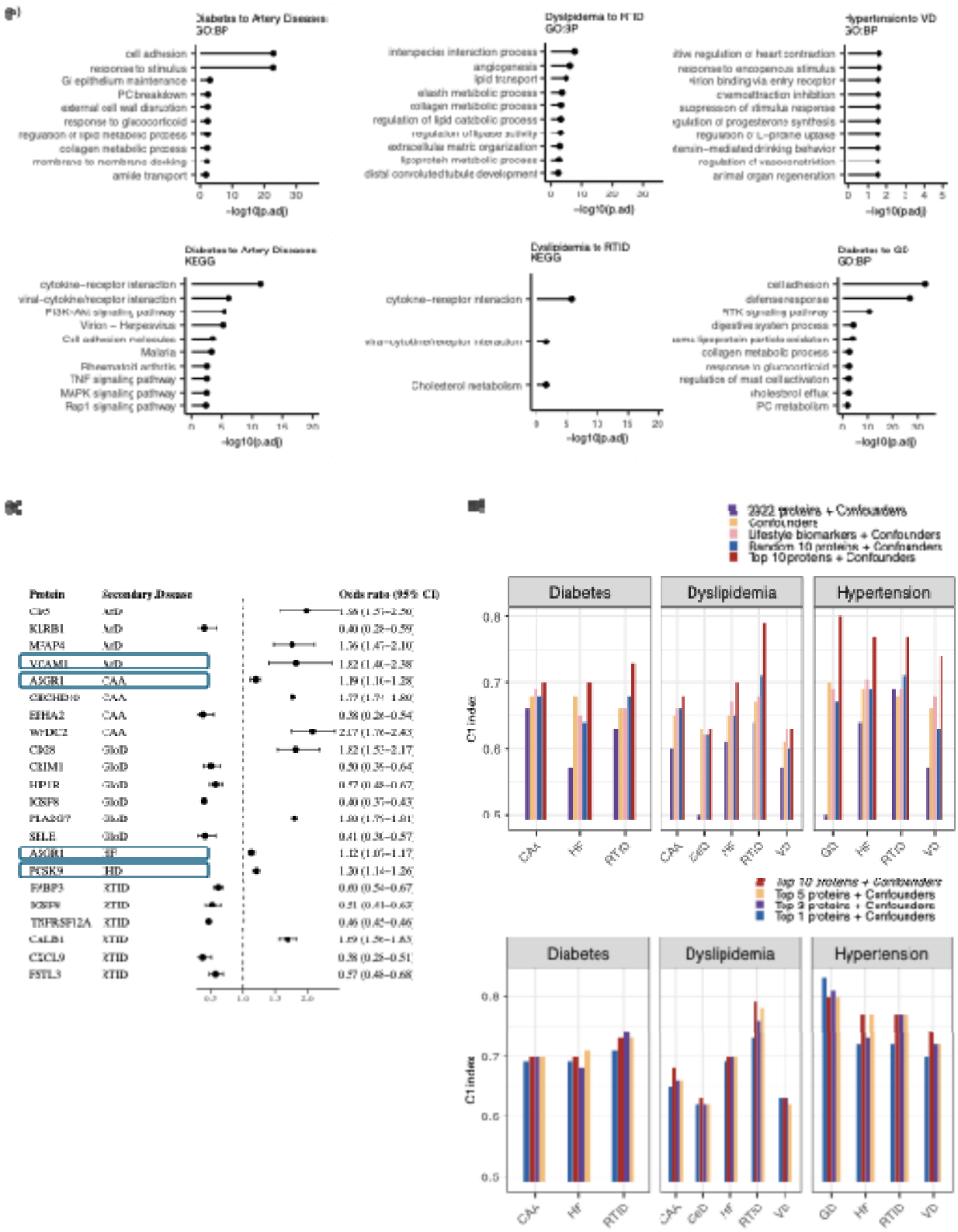


Fig. 3. Biological functional annotation and evaluation of identified protein mediators**.** (a) Pathway enrichment analysis of plasma proteins mediating selected disease paths (full results available in Supplementary Table). (b) Mendelian randomization results showing proteins with

potential causal effects on secondary diseases. (c) Comparison of prediction accuracy for secondary disease risk using machine learning models based on: (i) top 10 protein mediators combined with demographic confounders (age, sex, BMI), (ii) randomly selected sets of 10 proteins (100 repetitions), (iii) the full proteomic panel, and (iv) models built with varying numbers of top-ra nked proteins.

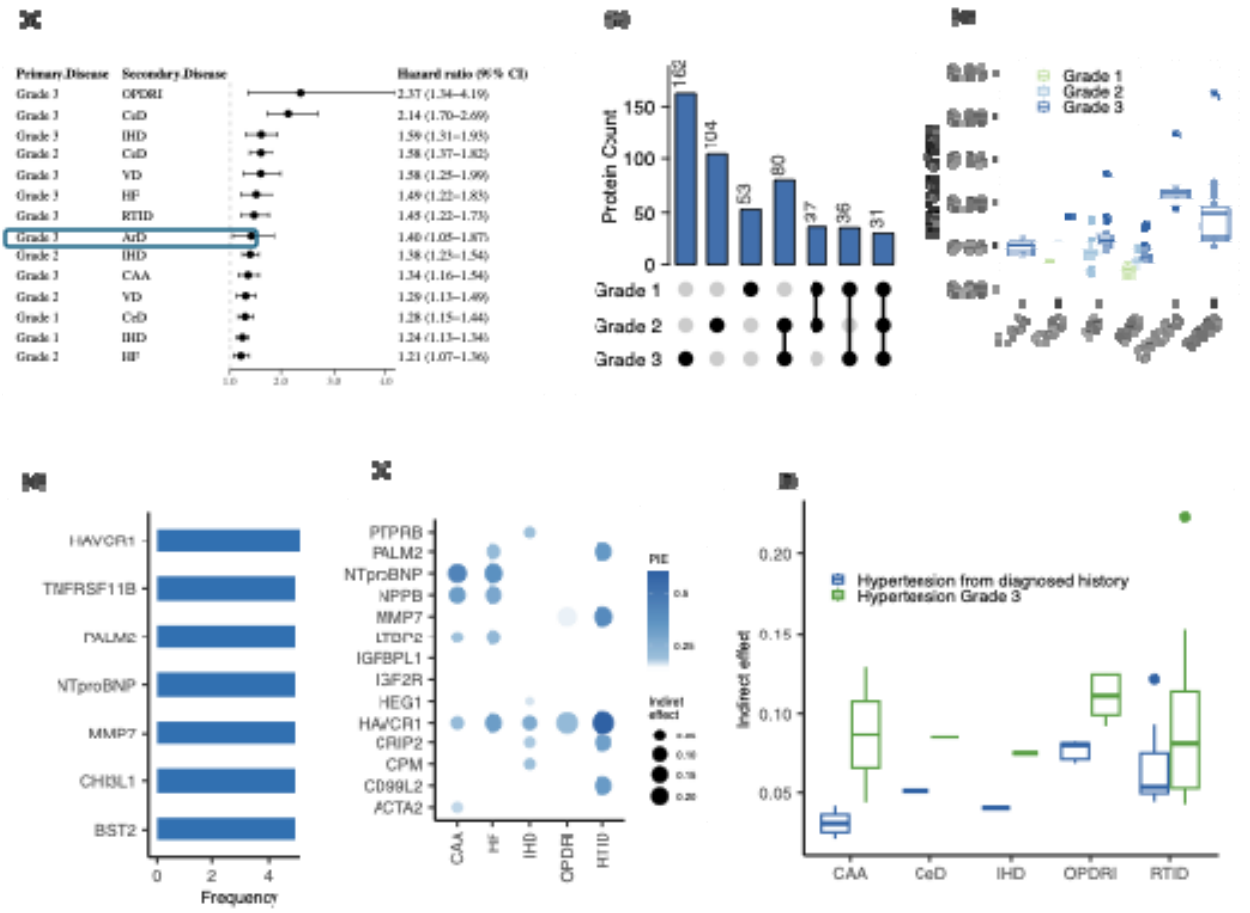


Fig. 4. Mediation analysis results using stratified hypertension severity grade based on blood pressure as exposure. (a) Hazard ratios and 95% confidence intervals for the associations between different hypertension grades and 18 secondary diseases; only statistically significant associations are shown. Notably, a new disease path, disease of the artery associated with grade 3 hypertension, was identified compared to binary hypertension diagnosis. (b) Number of plasma proteins mediating each grade-specific hypertension–disease path. (c) Comparison of mediation effects of proteins across different hypertension grades. (d) The plasma proteins frequently mediating disease paths linked to grade 3 hypertension. (e) Dot plot showing the top five plasma proteins with the strongest mediation effects in each disease path associated with grade 3 hypertension. (f) Comparison of mediation effects for the proteins linking to the same secondary disease, evaluated separately for grade 3 hypertension and binary hypertension diagnosis.

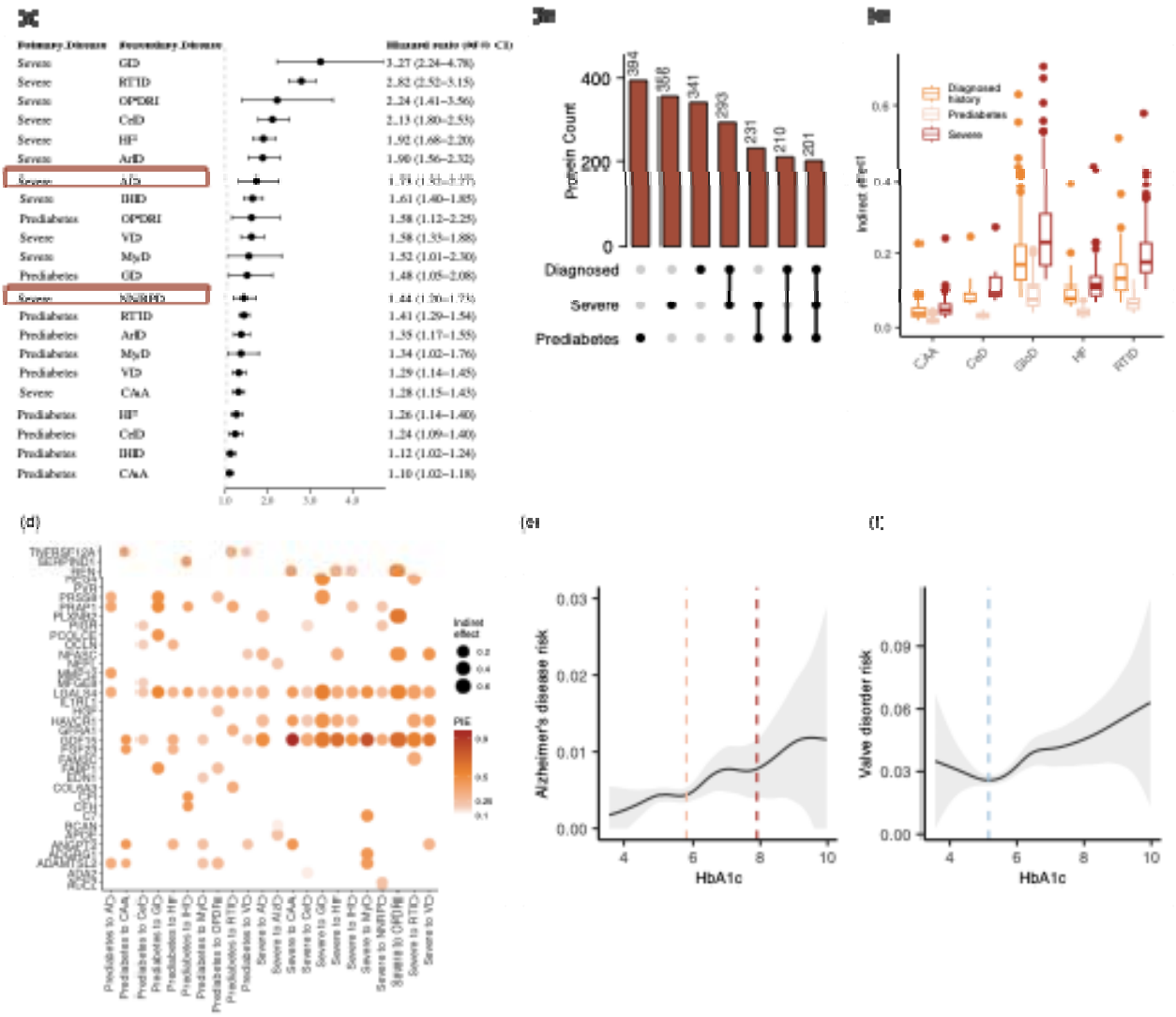


Fig. 5. Mediation analysis using diabetes severity grades stratified by HbA1c as the exposure. (a) Hazard ratios and 95% confidence intervals for the associations between different diabetes grades and 18 secondary diseases; only statistically significant associations are shown. (b) Number of plasma proteins mediating each diabetes-to-secondary disease pathway under different diabetes grades. (c) Comparison of mediation effects of proteins across different diabetes grades. (d) Dot plot showing the top five plasma proteins with the strongest mediation effects in each disease path associated with different diabetes grades. (e, f) Nonlinear associations between HbA1c levels and the risk of Alzheimer's disease (e) and valve disorders (f), estimated using generalized additive models. For Alzheimer's disease, the vertical lines indicate two inflection points beyond which disease risk increases rapidly. For valve disorders, the vertical line marks the HbA1c level associated with the lowest estimated risk.

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. Rosano, G. M. C., Vitale, C. & Seferovic, P. Heart failure in patients with diabetes mellitus. *Card Fail Rev* **3**, 52 (2017).
3. 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).
4. Matheus, A. S. de M. *et al.* Impact of diabetes on cardiovascular disease: an update. *Int J Hypertens* **2013**, 653789 (2013).
5. Resnick, H. E. & Howard, B. V. Diabetes and cardiovascular disease. *Annu Rev Med* **53**, 245–267 (2002).
6. Whelton, P. K. & Klag, M. J. Hypertension as a risk factor for renal disease. Review of clinical and epidemiological evidence. *Hypertension* **13**, I19 (1989).
7. Freitas Lima, L. C. *et al.* Adipokines, diabetes and atherosclerosis: an inflammatory association. *Front Physiol* **6**, (2015).
8. Savoia, C. & Schiffrin, E. L. Vascular inflammation in hypertension and diabetes: molecular mechanisms and therapeutic interventions. *Clin Sci* **112**, 375–384 (2007).
9. Granger, D. N., Vowinkel, T. & Petnehazy, T. Modulation of the Inflammatory Response in Cardiovascular Disease. *Hypertension* **43**, 924–931 (2004).
10. Assarsson, E. *et al.* Homogenous 96-Plex PEA Immunoassay Exhibiting High Sensitivity, Specificity, and Excellent Scalability. *PLoS One* **9**, e95192 (2014).
11. Gold, L. *et al.* Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery. *PLoS One* **5**, e15004 (2010).
12. Oh, H. S.-H. *et al.* Plasma proteomics links brain and immune system aging with healthspan and longevity. *Nat Med* (2025) doi:10.1038/s41591-025-03798-1.
13. Oh, H. S.-H. *et al.* Organ aging signatures in the plasma proteome track health and disease. *Nature* **624**, 164–172 (2023).
14. Carrasco-Zanini, J. *et al.* Proteomic signatures improve risk prediction for common and rare diseases. *Nat Med* **30**, 2489–2498 (2024).
15. Gadd, D. A. *et al.* Blood protein assessment of leading incident diseases and mortality in the UK Biobank. *Nat Aging* **4**, 939–948 (2024).
16. You, J. *et al.* Plasma proteomic profiles predict individual future health risk. *Nat Commun* **14**, 7817 (2023).
17. 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).
18. Deng, Y.-T. *et al.* Atlas of the plasma proteome in health and disease in 53,026 adults. *Cell* (2024) doi:10.1016/j.cell.2024.10.045.
19. Pearl, J. An Introduction to Causal Inference. *Int J Biostat* **6**, (2010).
20. 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).
21. Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).

22. Wanner, C. & Quaschning, T. Dyslipidemia and renal disease: pathogenesis and clinical consequences. *Curr Opin Nephrol Hypertens* **10**, 195–201 (2001).
23. 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).
24. 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).
25. Hedayatnia, M. *et al.* Dyslipidemia and cardiovascular disease risk among the MASHAD study population. *Lipids Health Dis* **19**, 42 (2020).
26. Sowers, J. R., Epstein, M. & Frohlich, E. D. Diabetes, Hypertension, and Cardiovascular Disease. *Hypertension* **37**, 1053–1059 (2001).
27. 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).
28. 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).
29. Seidah, N. G. PCSK9 as a therapeutic target of dyslipidemia. *Expert Opin Ther Targets* **13**, 19–28 (2009).
30. Wang, D. *et al.* GDF15: emerging biology and therapeutic applications for obesity and cardiometabolic disease. *Nat Rev Endocrinol* **17**, 592–607 (2021).
31. Li, J. *et al.* Overview of growth differentiation factor 15 (GDF15) in metabolic diseases. *Biomedicine & Pharmacotherapy* **176**, 116809 (2024).
32. 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).
33. Ishimitsu, T. *et al.* Genomic Structure of Human Adrenomedullin Gene. *Biochem Biophys Res Commun* **203**, 631–639 (1994).
34. 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).
35. Aleksander, S. A. *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**, (2023).
36. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat Genet* **25**, 25–29 (2000).
37. Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **28**, 27–30 (2000).
38. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Res* **9**, 709 (2020).
39. Richmond, R. C. & Davey Smith, G. Mendelian Randomization: Concepts and Scope. *Cold Spring Harb Perspect Med* **12**, a040501 (2022).
40. Elsworth, B. *et al.* The MRC IEU OpenGWAS data infrastructure. Preprint at https://doi.org/10.1101/2020.08.10.244293 (2020).
41. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, (2018).
42. Ley, K. & Huo, Y. VCAM-1 is critical in atherosclerosis. *Journal of Clinical Investigation* **107**, 1209–1210 (2001).

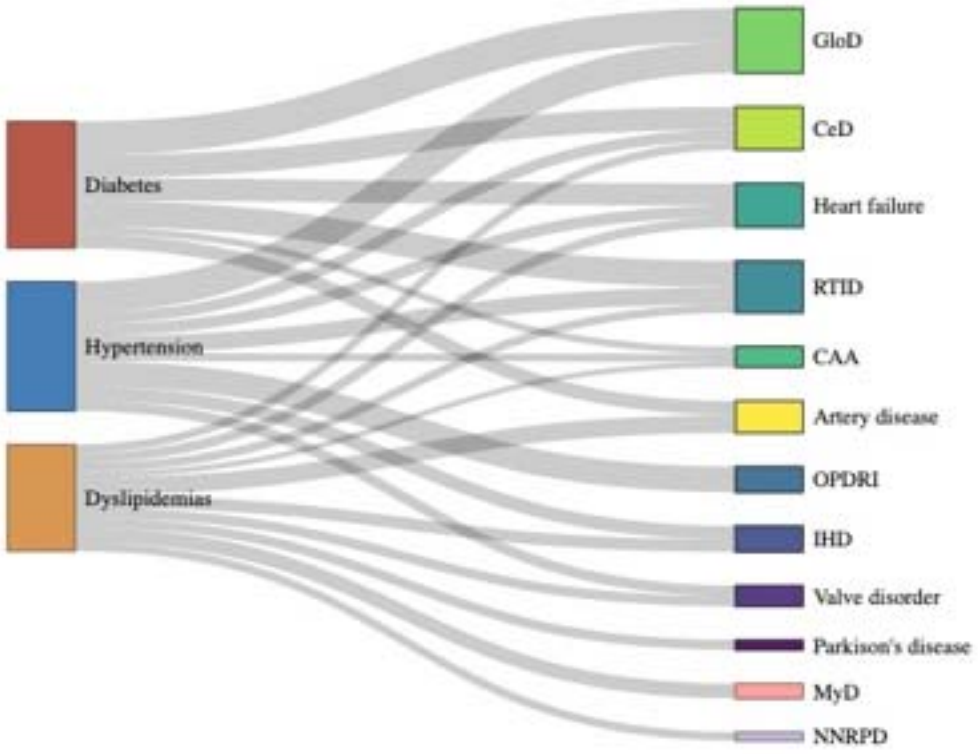
43. Sobati, S. *et al.* PCSK9: A Key Target for the Treatment of Cardiovascular Disease (CVD). *Advanced Pharmaceutical Bulletin* vol. 10 502–511 Preprint at https://doi.org/10.34172/apb.2020.062 (2020).
44. Mancia, G. *et al.* 2023 ESH Guidelines for the management of arterial hypertension The Task Force for the management of arterial hypertension of the European Society of Hypertension. *J Hypertens* **41**, 1874–2071 (2023).
45. Ozturk, T. C., Unluer, E., Denizbasi, A., Guneysel, O. & Onur, O. Can NT-proBNP be used as a criterion for heart failure hospitalization in emergency room? *J Res Med Sci* **16**, 1564 (2011).
46. ElSayed, N. A. *et al.* 2. Diagnosis and Classification of Diabetes: Standards of Care in Diabetes—2025. *Diabetes Care* **48**, S27–S49 (2025).
47. 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).
48. 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).
49. Wood, S. N. *Generalized Additive Models: An Introduction with R, Second Edition*. *Generalized Additive Models: An Introduction with R, Second Edition* (2017). doi:10.1201/9781315370279.
50. 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).
51. Dhindsa, R. S. *et al.* Rare variant associations with plasma protein levels in the UK Biobank. *Nature* **622**, 339–347 (2023).
52. Genest, J. Lipoprotein disorders and cardiovascular risk. *J Inherit Metab Dis* **26**, 267–287 (2003).
53. 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).
54. Kjeldsen, S. E. Hypertension and cardiovascular risk: General aspects. *Pharmacol Res* **129**, 95–99 (2018).
55. Fuchs, F. D. & Whelton, P. K. High Blood Pressure and Cardiovascular Disease. *Hypertension* **75**, 285–292 (2020).
56. MacKinnon, D. P., Fairchild, A. J. & Fritz, M. S. Mediation Analysis. *Annu Rev Psychol* **58**, 593–614 (2007).
57. Tibshirani, R. J. & Efron, B. An introduction to the bootstrap. *Monographs on statistics and applied probability* **57**, (1993).
58. 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).
59. Aleksander, S. A. *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**, (2023).
60. Kolberg, L. *et al.* g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res* **51**, W207–W212 (2023).
61. Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
62. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, (2018).

63. Elsworth, B. *et al.* The MRC IEU OpenGWAS data infrastructure. Preprint at https://doi.org/10.1101/2020.08.10.244293 (2020).
64. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet* **13**, e1007081 (2017).
65. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, (2018).
66. Hastie, T., Tibshirani, R., Friedman, J. H. & Friedman, J. H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. vol. 2 (Springer, 2009).
67. 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).

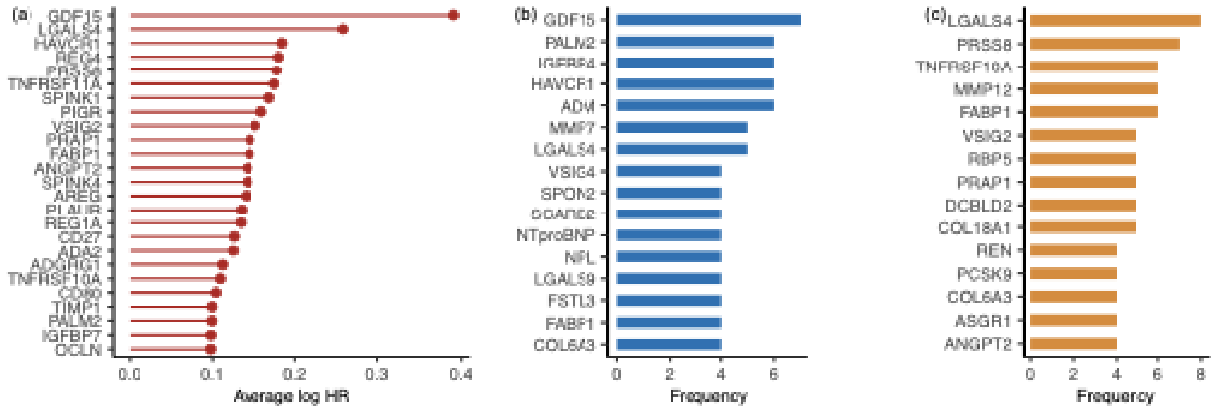
Acknowledgements
M.W., and H.Q., was supported by M.W.'s startup fund from the Institute for Heart and Brain Health. This work was supported by grants the NIH R35 HL155318 to A.R. and the American Heart Association (AHA) grants 23MERIT1038415 to A.R. and SFRN (URLs: https://doi.org/10.58275/ AHA.24SFRNPCN1284382.pc.gr.194135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

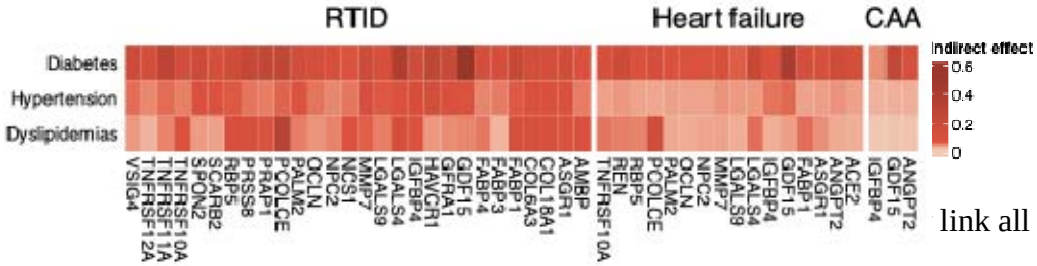
Competing interests
The authors declare no competing interests.



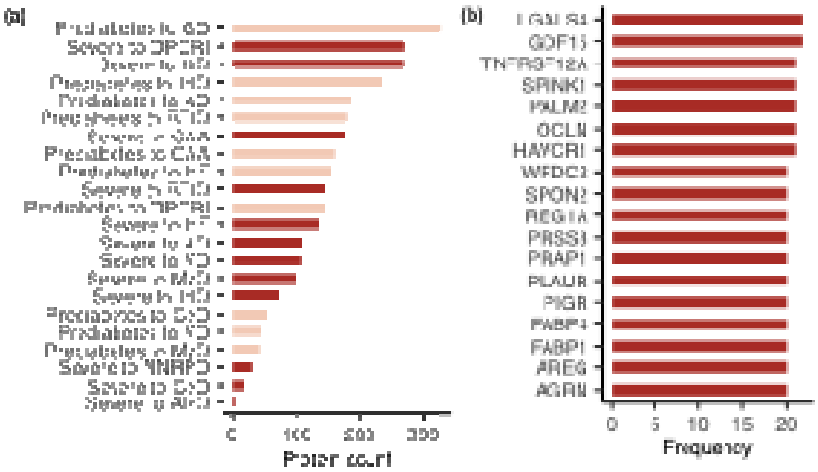
Supplementary Fig. 1. Diseases development pathways between three primary disease and secondary diseases in 4 clinical domain. The width of the line represent the hazard ratio.



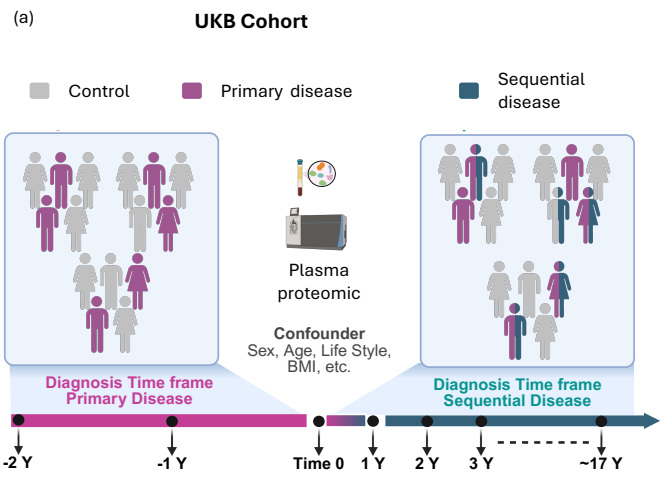
Supplementary Fig. 2. Important protein mediators identified based on the first criterion, shown separately for the three primary diseases: diabetes (a), hypertension (b), and dyslipidemia (c), respectively.

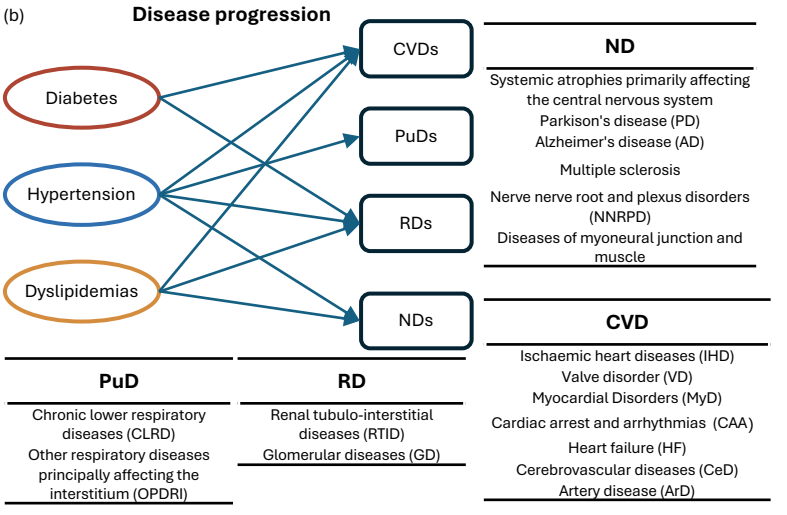


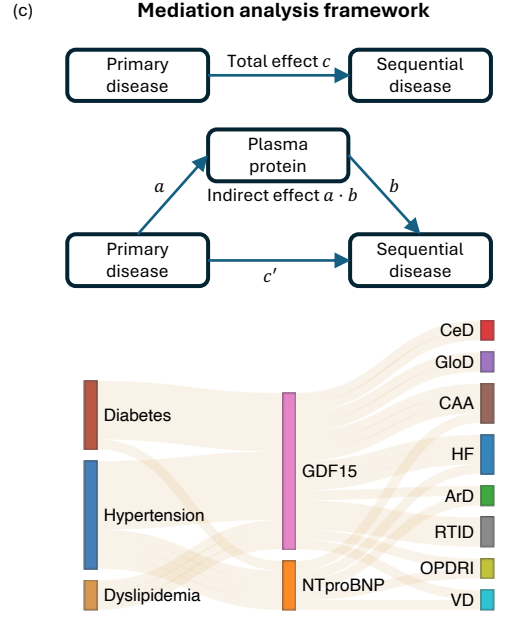
S link all three primary diseases to one secondary disease

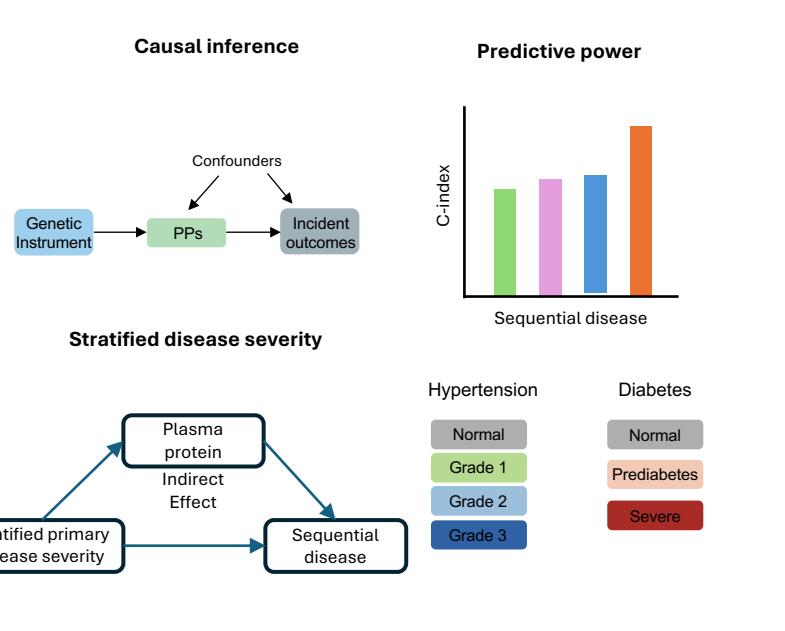


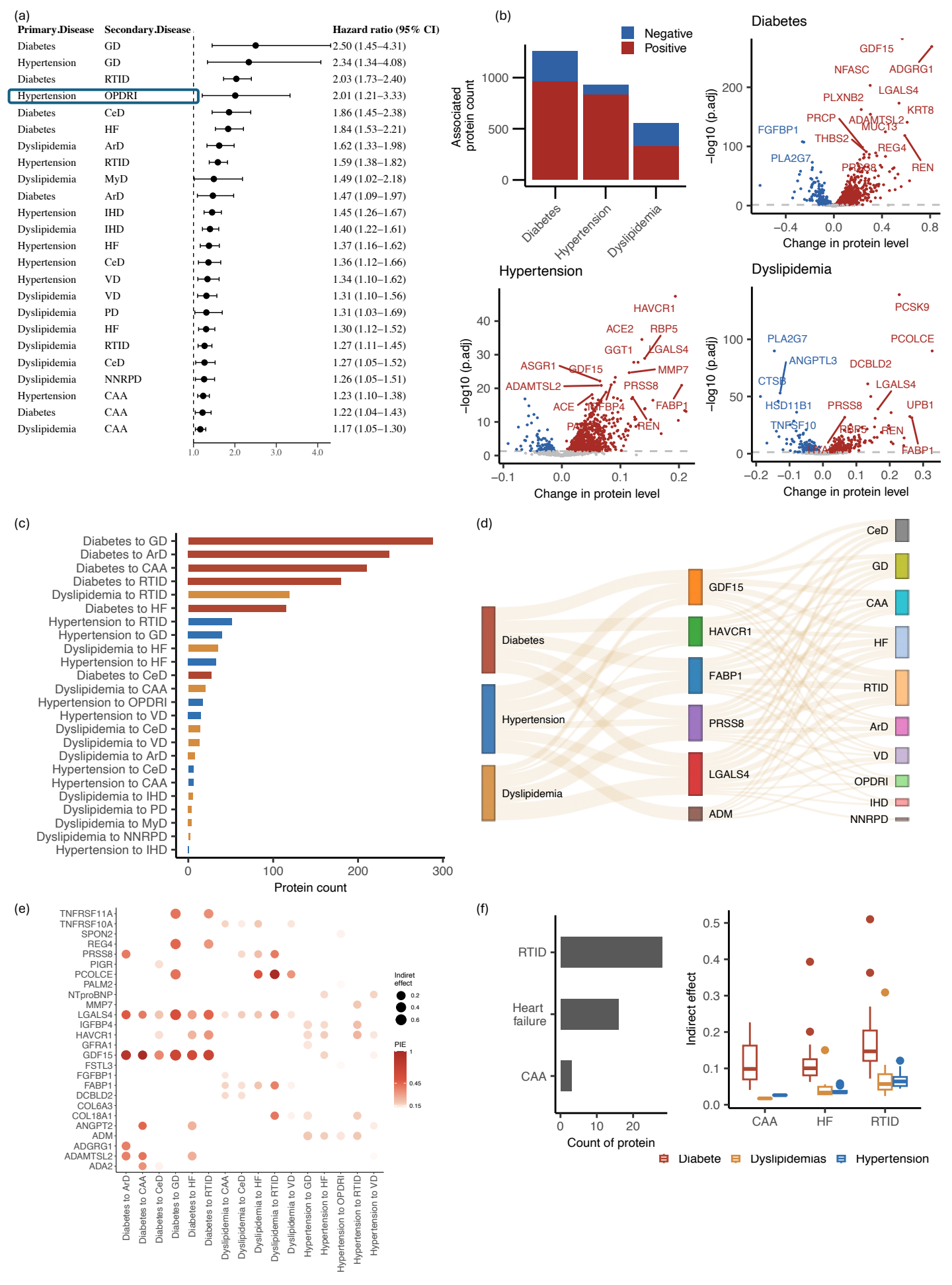
Supplementary Fig 4. Mediation analysis results stratified by diabetes severity levels based on HbA1c. (a) Number of plasma proteins mediating each disease path across different diabetes severity levels. (b) Frequently observed protein mediators that consistently link varying levels of diabetes severity to diverse secondary diseases.


(a)
UKB Cohort
Control
Primary disease
Sequential disease
Plasma proteomic
Confounder
Sex, Age, Life Style, BMI, etc.
Diagnosis Time frame Primary Disease
Diagnosis Time frame Sequential Disease
-2 Y
-1 Y
Time 0
1 Y
2 Y
3 Y
~17 Y

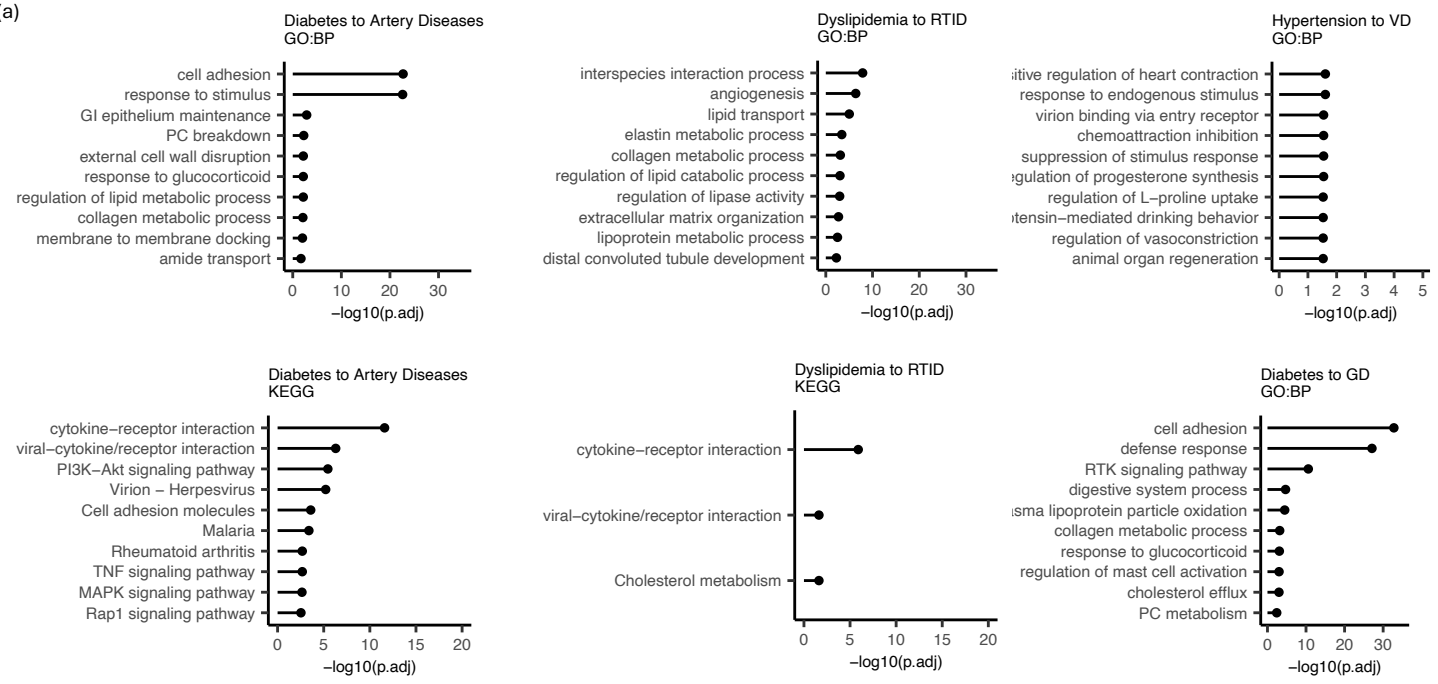

(b)
Disease progression
Diabetes
Hypertension
Dyslipidemias
CVDs
PuDs
RDs
NDs
ND
Systemic atrophies primarily affecting the central nervous system
Parkison's disease (PD)
Alzheimer's disease (AD)
Multiple sclerosis
Nerve nerve root and plexus disorders (NNRPD)
Diseases of myoneural junction and muscle
CVD
Ischaemic heart diseases (IHD)
Valve disorder (VD)
Myocardial Disorders (MyD)
Cardiac arrest and arrhythmias (CAA)
Heart failure (HF)
Cerebrovascular diseases (CeD)
Artery disease (ArD)
PuD
Chronic lower respiratory diseases (CLRD)
Other respiratory diseases principally affecting the interstitium (OPDRI)
RD
Renal tubulo-interstitial diseases (RTID)
Glomerular diseases (GD)


(c)
Mediation analysis framework
Primary disease
Total effect c
Sequential disease
Plasma protein
a
b
Indirect effect a · b
Primary disease
c′
Sequential disease
Diabetes
Hypertension
Dyslipidemia
GDF15
NTproBNP
CeD
GloD
CAA
HF
ArD
RTID
OPDRI
VD

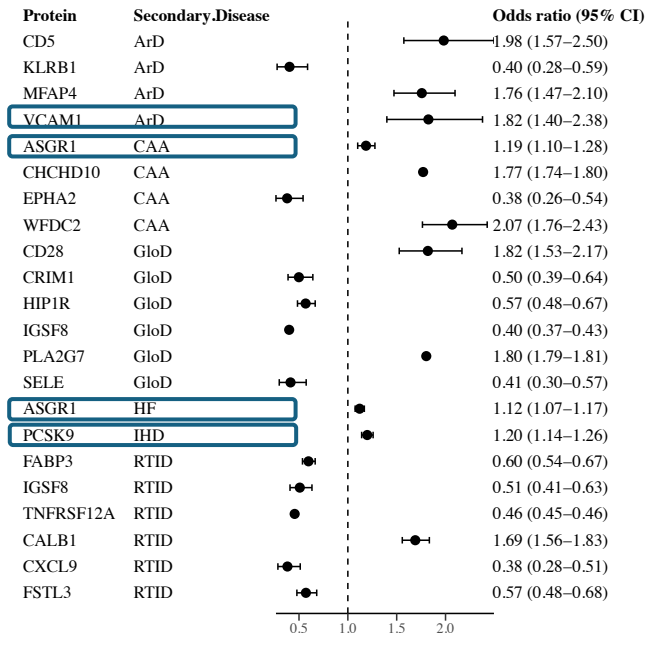

(d)
Causal inference
Confounders
Genetic Instrument
PPs
Incident outcomes
Predictive power
C-index
Sequential disease
(e)
Stratified disease severity
Plasma protein
Indirect Effect
Stratified primary disease severity
Sequential disease
Hypertension
Normal
Grade 1
Grade 2
Grade 3
Diabetes
Normal
Prediabetes
Severe


(a)
Primary.Disease
Secondary.Disease
Hazard ratio (95% CI)
Diabetes GD 2.50 (1.45–4.31)
Hypertension GD 2.34 (1.34–4.08)
Diabetes RTID 2.03 (1.73–2.40)
Hypertension OPDRI 2.01 (1.21–3.33)
Diabetes CeD 1.86 (1.45–2.38)
Diabetes HF 1.84 (1.53–2.21)
Dyslipidemia ArD 1.62 (1.33–1.98)
Hypertension RTID 1.59 (1.38–1.82)
Dyslipidemia MyD 1.49 (1.02–2.18)
Diabetes ArD 1.47 (1.09–1.97)
Hypertension IHD 1.45 (1.26–1.67)
Dyslipidemia IHD 1.40 (1.22–1.61)
Hypertension HF 1.37 (1.16–1.62)
Hypertension CeD 1.36 (1.12–1.66)
Hypertension VD 1.34 (1.10–1.62)
Dyslipidemia VD 1.31 (1.10–1.56)
Dyslipidemia PD 1.31 (1.03–1.69)
Dyslipidemia HF 1.30 (1.12–1.52)
Dyslipidemia RTID 1.27 (1.11–1.45)
Dyslipidemia CeD 1.27 (1.05–1.52)
Dyslipidemia NNRPD 1.26 (1.05–1.51)
Hypertension CAA 1.23 (1.10–1.38)
Diabetes CAA 1.22 (1.04–1.43)
Dyslipidemia CAA 1.17 (1.05–1.30)
1.0 2.0 3.0 4.0
(b)
Negative
Positive
Associated protein count
Diabetes
Hypertension
Dyslipidemia
Diabetes
−log10 (p.adj)
Change in protein level
GDF15
NFASC
ADGRG1
LGALS4
PLXNB2
KRT8
PRCP
ADAMTSL2
MUC13
FGFBP1
THBS2
REG4
PLA2G7
REN
Hypertension
HAVCR1
ACE2
RBP5
GGT1
LGALS4
ASGR1
GDF15
MMP7
ADAMTSL2
PRSS8
ACE
FABP4
FABP1
REN
Dyslipidemia
PCSK9
PLA2G7
PCOLCE
ANGPTL3
DCBLD2
CTSB
LGALS4
HSD11B1
PRSS8
UPB1
TNFSF10
REN
FABP1
(c)
Diabetes to GD
Diabetes to ArD
Diabetes to CAA
Diabetes to RTID
Dyslipidemia to RTID
Diabetes to HF
Hypertension to RTID
Hypertension to GD
Dyslipidemia to HF
Hypertension to HF
Diabetes to CeD
Dyslipidemia to CAA
Hypertension to OPDRI
Hypertension to VD
Dyslipidemia to CeD
Dyslipidemia to VD
Dyslipidemia to ArD
Hypertension to CeD
Hypertension to CAA
Dyslipidemia to IHD
Dyslipidemia to PD
Dyslipidemia to MyD
Dyslipidemia to NNRPD
Hypertension to IHD
Protein count
(d)
Diabetes
Hypertension
Dyslipidemia
GDF15
HAVCR1
FABP1
PRSS8
LGALS4
ADM
CeD
GD
CAA
HF
RTID
ArD
VD
OPDRI
IHD
NNRPD
(e)
TNFRSF11A
TNFRSF10A
SPON2
REG4
PRSS8
PIGR
PCOLCE
PALM2
NTproBNP
MMP7
LGALS4
IGFBP4
HAVCR1
GFRA1
GDF15
FSTL3
FGFBP1
FABP1
DCBLD2
COL6A3
COL18A1
ANGPT2
ADM
ADGRG1
ADAMTSL2
ADA2
Diabetes to ArD
Diabetes to CAA
Diabetes to CeD
Diabetes to GD
Diabetes to HF
Diabetes to RTID
Dyslipidemia to CAA
Dyslipidemia to CeD
Dyslipidemia to HF
Dyslipidemia to RTID
Dyslipidemia to VD
Hypertension to GD
Hypertension to HF
Hypertension to OPDRI
Hypertension to RTID
Hypertension to VD
Indiret effect
PIE
(f)
RTID
Heart failure
CAA
Count of protein
Indirect effect
CAA
HF
RTID
Diabete
Dyslipidemias
Hypertension

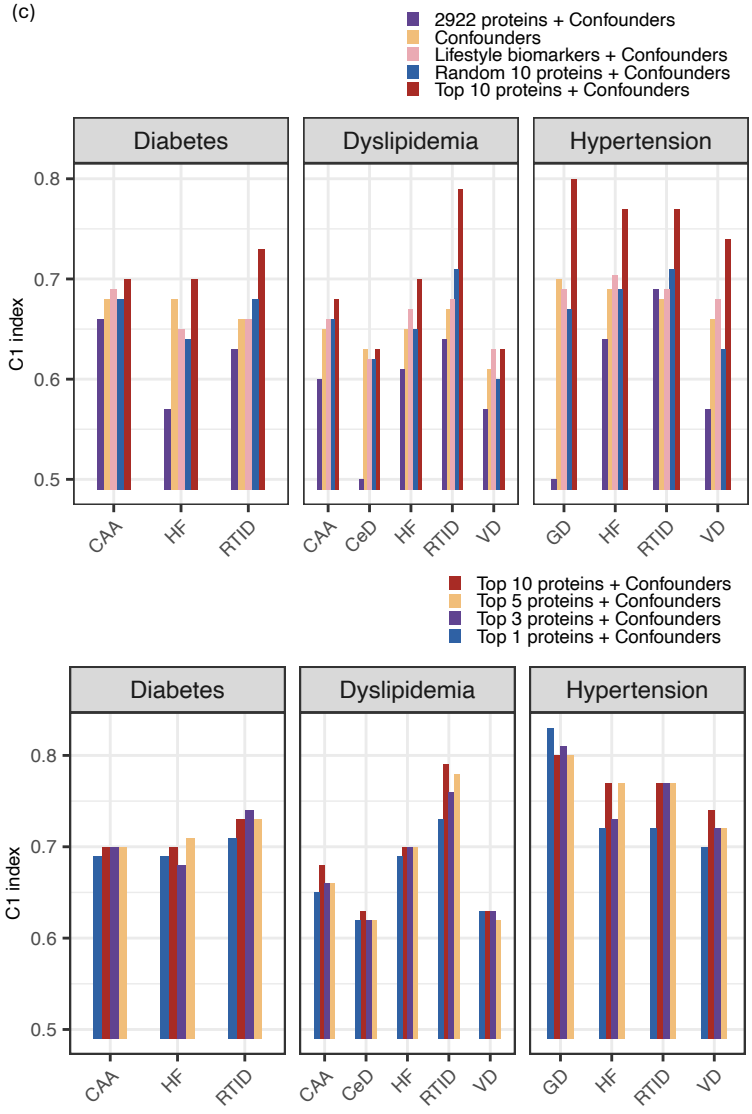
(a)

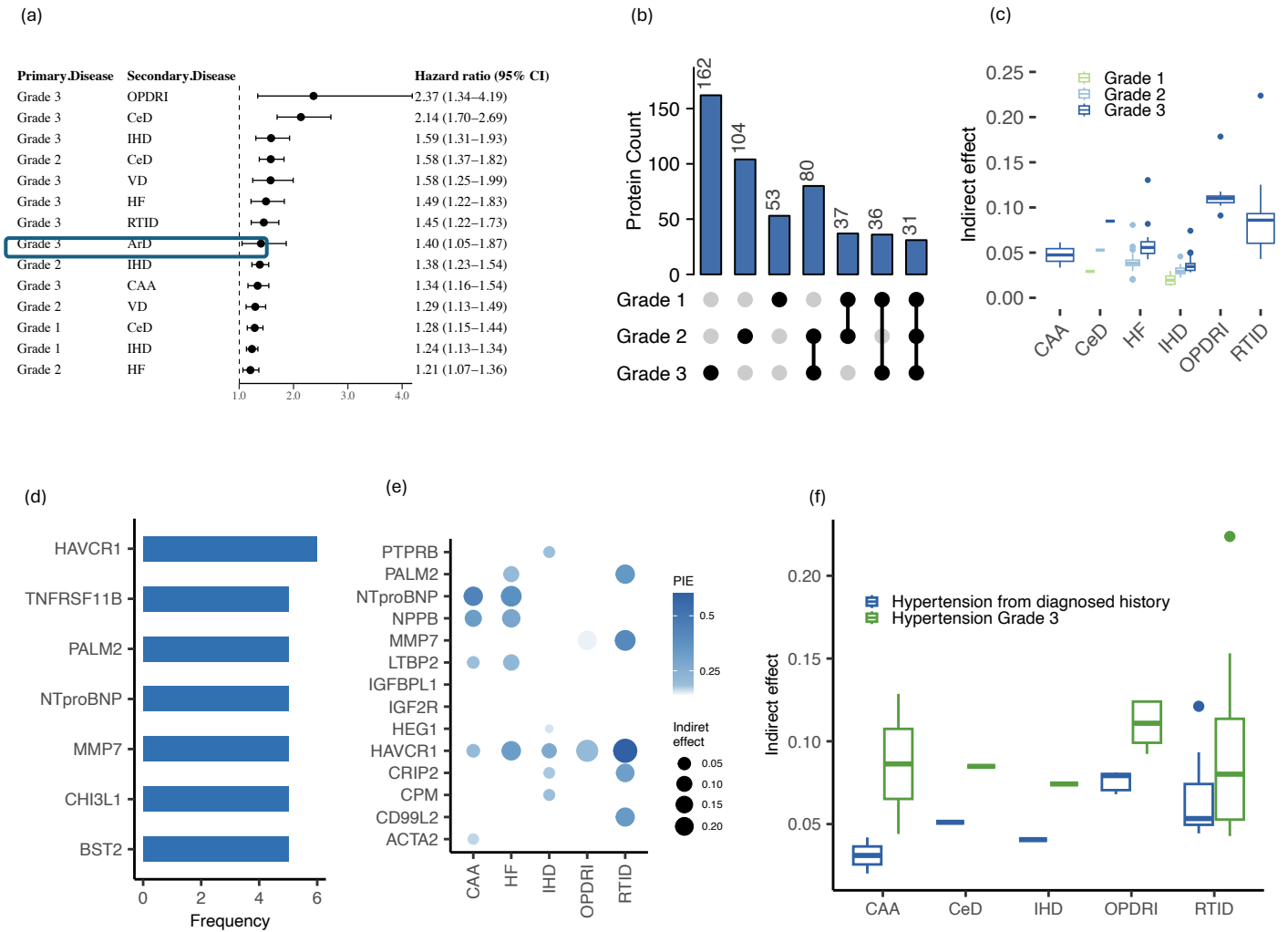

Diabetes to Artery Diseases
GO:BP
cell adhesion
response to stimulus
GI epithelium maintenance
PC breakdown
external cell wall disruption
response to glucocorticoid
regulation of lipid metabolic process
collagen metabolic process
membrane to membrane docking
amide transport
0 10 20 30
−log10(p.adj)
Dyslipidemia to RTID
GO:BP
interspecies interaction process
angiogenesis
lipid transport
elastin metabolic process
collagen metabolic process
regulation of lipid catabolic process
regulation of lipase activity
extracellular matrix organization
lipoprotein metabolic process
distal convoluted tubule development
0 10 20 30
−log10(p.adj)
Hypertension to VD
GO:BP
itive regulation of heart contraction
response to endogenous stimulus
virion binding via entry receptor
chemoattraction inhibition
suppression of stimulus response
gulation of progesterone synthesis
regulation of L−proline uptake
tensin−mediated drinking behavior
regulation of vasoconstriction
animal organ regeneration
0 1 2 3 4 5
−log10(p.adj)
Diabetes to Artery Diseases
KEGG
cytokine−receptor interaction
viral−cytokine/receptor interaction
PI3K−Akt signaling pathway
Virion − Herpesvirus
Cell adhesion molecules
Malaria
Rheumatoid arthritis
TNF signaling pathway
MAPK signaling pathway
Rap1 signaling pathway
0 5 10 15 20
−log10(p.adj)
Dyslipidemia to RTID
KEGG
cytokine−receptor interaction
viral−cytokine/receptor interaction
Cholesterol metabolism
0 5 10 15 20
−log10(p.adj)
Diabetes to GD
GO:BP
cell adhesion
defense response
RTK signaling pathway
digestive system process
sma lipoprotein particle oxidation
collagen metabolic process
response to glucocorticoid
regulation of mast cell activation
cholesterol efflux
PC metabolism
0 10 20 30
−log10(p.adj)

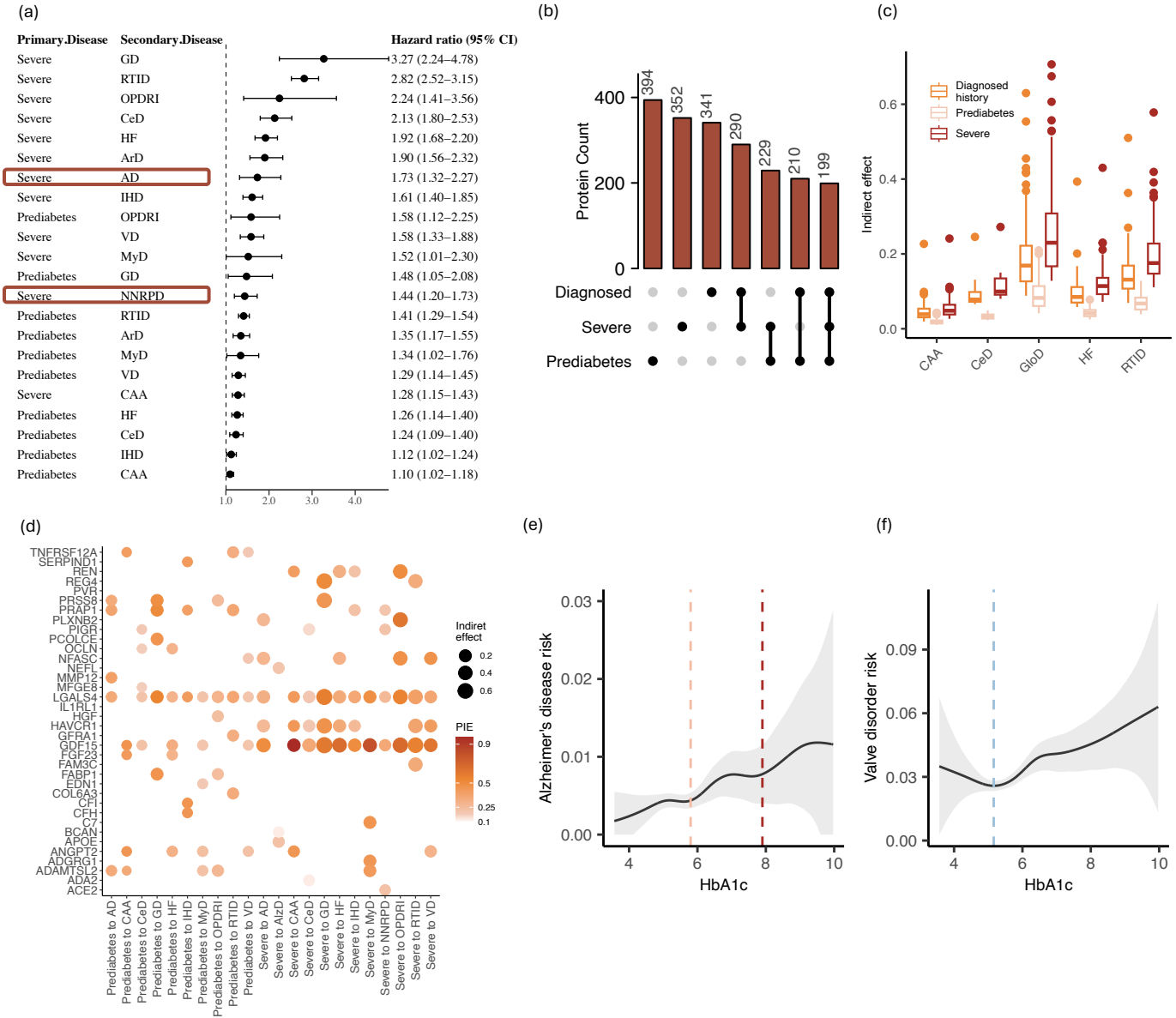
(b)


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 GloD 1.82 (1.53–2.17)
CRIM1 GloD 0.50 (0.39–0.64)
HIP1R GloD 0.57 (0.48–0.67)
IGSF8 GloD 0.40 (0.37–0.43)
PLA2G7 GloD 1.80 (1.79–1.81)
SELE GloD 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)


2922 proteins + Confounders
Confounders
Lifestyle biomarkers + Confounders
Random 10 proteins + Confounders
Top 10 proteins + Confounders
Diabetes
Dyslipidemia
Hypertension
C1 index
0.8
0.7
0.6
0.5
CAA HF RTID
CAA CeD HF RTID VD
GD HF RTID VD
Top 10 proteins + Confounders
Top 5 proteins + Confounders
Top 3 proteins + Confounders
Top 1 proteins + Confounders
Diabetes
Dyslipidemia
Hypertension
C1 index
0.8
0.7
0.6
0.5
CAA HF RTID
CAA CeD HF RTID VD
GD HF RTID VD


(a)
Primary.Disease
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
(b)
Protein Count
162
104
53
80
37
36
31
Grade 1
Grade 2
Grade 3
(c)
Indirect effect
Grade 1
Grade 2
Grade 3
CAA
CeD
HF
IHD
OPDRI
RTID
(d)
HAVCR1
TNFRSF11B
PALM2
NTproBNP
MMP7
CHI3L1
BST2
Frequency
(e)
PTPRB
PALM2
NTproBNP
NPPB
MMP7
LTBP2
IGFBPL1
IGF2R
HEG1
HAVCR1
CRIP2
CPM
CD99L2
ACTA2
CAA
HF
IHD
OPDRI
RTID
PIE
Indiret effect
(f)
Hypertension from diagnosed history
Hypertension Grade 3
Indirect effect
CAA
CeD
IHD
OPDRI
RTID


(a)
Primary.Disease
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)
1.0
2.0
3.0
4.0
(b)
Protein Count
394
352
341
290
229
210
199
Diagnosed
Severe
Prediabetes
(c)
Indirect effect
Diagnosed history
Prediabetes
Severe
CAA
CeD
GloD
HF
RTID
(d)
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
(e)
Alzheimer's disease risk
HbA1c
(f)
Valve disorder risk
HbA1c