Plasma proteins and onset of type 2 diabetes and diabetic complications: Proteome-wide Mendelian randomization and colocalization analyses

Highlights
- A proteome-wide analysis has been conducted for type 2 diabetes and complications
- Forty-seven plasma proteins are causally associated with type 2 diabetes
- Seventeen proteins are associated causally with at least one diabetic complication
- HLA-DRA, AGER, HSPA1A, and HSPA1B are associated with microvascular complications

Authors
Shuai Yuan, Fengzhe Xu, Xue Li, Jie Chen, Jie Zheng, Christos S. Mantzoros, Susanna C. Larsson

Correspondence
shuai.yuan@ki.se (S.Y.), susanna.larsson@ki.se (S.C.L.)

In brief
To determine the causal associations of blood proteins with type 2 diabetes and its complications, Yuan et al. perform a proteome-wide MR investigation and identify many circulating proteins in relation to type 2 diabetes and four proteins with universal effects on microvascular complications.
We conduct proteome-wide Mendelian randomization and colocalization analyses to decipher the associations of blood proteins with the risk of type 2 diabetes and diabetic complications. Genetic data on plasma proteome are obtained from 54,306 UK Biobank participants and 35,559 Icelanders. Summary-level data on type 2 diabetes are obtained from the DIAGRAM (DIAbetes Genetics Replication And Meta-analysis consortium) consortium (74,124 cases) and FinnGen study (33,043 cases). Data on 10 diabetic complications are obtained from FinnGen and corresponding studies. Among 1,886 proteins, genetically predicted levels of 47 plasma proteins are associated with type 2 diabetes. Eleven of these proteins have strong support of colocalization. Seventeen proteins are associated with at least one diabetic complication, although a few have colocalization support. HLA-DRA, AGER, HSPA1A, and HSPA1B are associated with most microvascular complications. This study reveals causal proteins for the onset of type 2 diabetes and diabetic complications, which enhances the understanding of molecular etiology and development of therapeutics.
Diabetic complications lead to a heavy disease burden, such as blindness, kidney dysfunction, an impaired quality of life, and premature mortality among those with diabetes. Several drugs, like metformin, sodium-glucose co-transporter-2 (SGLT2) inhibitors, and glucagon-like peptide-1 receptor agonists, have been used to lower the risk of developing complications or delay their onset among diabetic patients. An appraisal of shared proteomic basis between diabetes and diabetic complications is also of importance to better understand underlying pathophysiology and possibly to manage disease progression. It was found in a cohort of 528 individuals that 15 plasma proteins shared by type 2 diabetes and coronary artery disease (a major macrovascular complication of diabetes). However, the shared proteomic etiology between diabetes and diabetic microvascular complications remains to be fully explored.

Proteomic research can not only deepen molecular understanding, but also help to reveal potential therapeutic targets, since many circulating proteins always act as the principal regulators of molecular pathways. For example, some classes of drugs used for type 2 diabetes target proteins, like alpha-glucosidase inhibitors that inhibit alpha-amylase and alpha-glucosidase, the key enzymes involved in the digestion of carbohydrates, and SGLT2 inhibitors that inhibit sodium-glucose transport proteins in the nephron. MR design is an epidemiological approach that can strengthen causal inference by using genetic variants as an instrumental variable for the exposure (e.g., circulating levels of a protein). There are three key assumptions of MR: (1) the genetic variants used as the instrumental variable should be robustly associated with the exposure of interest; (2) the genetic instruments should not be associated with any confounders; and (3) the genetic instruments should impact the outcome only via the exposure of interest and not via alternative pathways. Using MR analysis to explore the associations between plasma proteins and health outcomes has been widely used and found to be assumption satisfied when using the cis-variant in the protein-encoding gene as the genetic instrument for the protein. Some previous MR studies have been conducted to estimate the association between blood proteins and the risk of type 2 diabetes. Based on a larger number of blood proteins with data from two large-scale studies, we conducted a protein-wide MR study supplemented by colocalization analysis to explore nearly 2,000 plasma proteins in relation to the onset of type 2 diabetes and diabetic complications.

**RESULTS**

An overview of the study design is shown in Figure 1. All analyses were based on summary-level data listed in Table 1. We replaced 132 missing SNPs with proxy SNPs in FinnGen. After removing plasma proteins without genetic instruments or SNP proxies in type 2 diabetes data, MR analysis included 1,797 proteins in the DIAGRAM (DIAbetes Genetics Replication And Meta-analysis consortium) consortium, 1,835 proteins in the FinnGen R8 study, and 1,885 proteins in the combined analysis. The minimal F statistic of used genetic instruments was more than 300 in all outcome data. The statistical power in the analyses of diabetic complications was generally low because of small sample sizes (Figure S1).

**Associations between plasma proteins and type 2 diabetes**

Figure 2 displays the result summary of the analyses of type 2 diabetes. In the combined analysis of two outcomes data, genetically predicted levels of 47 proteins were significantly associated with the risk of type 2 diabetes after Bonferroni correction for multiple testing ($p < 2.65 \times 10^{-8}$) (Figure 2A). Per SD increase in genetically predicted levels of protein, the odds ratio of type 2 diabetes ranged from 0.38 (95% confidence interval [CI], 0.28–0.51) for heat shock 70 kDa protein 1B (HSPA1B) to 2.08 (95% CI, 1.84–2.34) for glucokinase regulatory protein (GCKR) (Figure 2B and Table S1). The identified associations were directionally consistent between the discovery and replication studies for all proteins except heat shock 70 kDa protein 1A (HSPA1A) (Tables S2 and S3). Among identified associations, eight proteins were measured in both outcome datasets and generally showed consistent associations with type 2 diabetes, despite the proteins being measured on two different profiling platforms (Table S4). Genetically predicted levels of the other studied proteins were not associated with the risk of type 2 diabetes (Table S1).

Among 47 MR-identified proteins in relation to type 2 diabetes, 9 proteins had high support of colocalization analysis ($PH4 \geq 0.8$) (Figure 2C and Table S5), which were GCKR, histo-blood group ABO system transferase (ABO), hedgehog-interacting protein (HHIP), phosphoglucomutase-1 (PGM1), mitogen-activated protein kinase 8 (MAPK8), metallophosphoesterase domain-containing protein 2 (MPPED2), 3-hydroxyisobutyryl-CoA dehydratase, mitochondrial (HIBCH), glutathione S-transferase A1 (GSTA1), and ERO1-like protein beta. Three proteins had medium support of colocalization analysis ($0.8 \geq PH4 \geq 0.5$) (Figure 2C). Sum of single effects (SuSiE) additionally identified strong colocalization support for the associations for arginase 1 (ARG1) and probable tRNA(His) guanylyltransferase (THG1L) (Figure 2C).

**Function and network prediction of diabetes-associated proteins**

Identified diabetes-associated proteins had networks, in particular co-expression and physical interactions (Figure S2). Based on cis-genes for these proteins, many pathways were enriched, including protein-folding and catabolism, response to temperature stimulus/heat, intrinsic apoptotic signaling pathway, and response to oxidative stress (Table S6).

**Reverse MR analysis of type 2 diabetes liability with levels of identified 47 proteins**

Genetic liability to type 2 diabetes was associated with levels of six blood proteins (THG1L, sex hormone-binding globulin [SHBG]), inhibit beta B chain, chymotrypsinogen B2, major histocompatibility complex class I polypeptide-related sequence B [MICB], and HIBCH) after Bonferroni multiple testing correction. The associations were consistent in a series of analyses (Table S7).

**Associations between diabetes-associated proteins and body mass index**

Among 47 diabetes-associated proteins, genetically predicted levels of 18 proteins were associated with body mass index (BMI) in the GIANT (The Genetic Investigation of ANthropometric...
The associations were overall consistent between the two sources. Per SD increase, the changes of BMI ranged from –0.29 (95% CI, –0.37 to –0.20) for mitochondrial peptide methionine sulfoxide reductase to 0.12 (95% CI, 0.08–0.16) for GCKR. The two-stage network MR analysis identified pathways mediated by BMI from protein to type 2 diabetes. The mediation by BMI ranged from 15.8% for the GCKR-diabetes association to 60.6% for the hypoxia up-regulated protein 1-diabetes association (Table S8). Mainly because of the high co-expression of these proteins (Figure S4), the sum of mediation proportion was greater than 100%.

Disrupted metabolic pathways and their associations with diabetic complications

Figure 3 displays the result summary of the analyses of 47 diabetes-associated plasma proteins in relation to major diabetic complications. After adjusting for multiple testing, genetically predicted levels of 17 proteins were associated with at least one diabetic complication in MR analysis (Figures 3A and 3B). The directions of the protein-diabetes associations were overall consistent with that of the associations between proteins and diabetic complications (Figure 3A), except for genetically predicted levels of PGM1 with an inverse association with type 2 diabetes, but a positive association with diabetic ketoacidosis (Figure 3A). Diabetes-associated proteins were more likely to be associated with acute and chronic microvascular complications instead of macrovascular complications (Figure 3A). Four of these proteins that included human leukocyte antigen class II histocompatibility antigen, DR alpha chain (HLA-DRA), advanced glycosylation end-product-specific receptor (AGER), HSPA1A, and HSPA1B were directionally consistently associated with most studied diabetic complications (Figure 3B). Genetically predicted levels of the other proteins were not associated with any of studied diabetic complications (Table S9).

Some of above MR associations were supported by traditional and SuSiE colocalization analyses with PH4 >0.5 (Figure 3C and Table S10). MANSC domain-containing protein 4 (MANSC4) had moderate to strong support of colocalization with diabetic ketoacidosis (PH4 = 0.92), diabetic hypoglycemia (PH4 = 0.67 in traditional and 0.96 in SuSiE), and diabetic retinopathy (PH4 = 0.63 in traditional and 0.85 in SuSiE). Beta-mannosidase (MANBA) had moderate support of colocalization with diabetic retinopathy (PH4 = 0.65 in traditional). PGM1 had moderate support of colocalization with diabetic ketoacidosis (PH4 = 0.59 in traditional). PTPN9 had moderate support of colocalization with diabetic maculopathy (PH4 = 0.63 in traditional).

**DISCUSSION**

This MR study examined the associations of 1,885 plasma proteins with the risk of type 2 diabetes and was supplemented by a colocalization analysis. We found 47 plasma proteins with a potential causal association with type 2 diabetes, of which 11 proteins had a strong support of colocalization. Function prediction based on these proteins enriched many pathways, including protein folding and catabolism, response to temperature stimulus/heat, intrinsic apoptotic signaling pathway, and response to oxidative stress. Several of these proteins were associated with BMI in a consistent direction with that for diabetes and the two-stage network MR indicated partial mediation of BMI in these protein-diabetes associations. The subsequent analysis revealed the roles of 17 diabetes-associated proteins in diabetic complications, although only some proteins were identified in colocalization analysis. These diabetes-associated proteins were more likely to be associated with microvascular complications compared with macrovascular complications. This analysis further found a pervasive effect of four proteins (i.e., HLA-DRA, AGER, HSPA1A, and HSPA1B) on acute and chronic microvascular complications. Our study corroborated some previously identified protein-diabetes associations, such as the associations of diabetes...
with TYRO3 (tyrosine-protein kinase receptor TYRO3), ARG1, SHBG, MANSC4, HP, MANBA, ABO, angiotensin-converting enzyme (ACE), ERO1-like protein beta (ERO1LB), MICB, and GSTA1.8,14 The associations of some well-studied proteins, like GCKR,5 SHBG,16 ABO,17 and ACE18 with the risk of developing type 2 diabetes were also identified, which indicated a good validity of data sources used in the current analysis. However, except for ABO, ARG1, ERO1LB, GSTA1, and TYRO3, we observed weak colocalization support of the associations for other above-mentioned proteins, which implies that these associations may be influenced by linkage disequilibrium. However, a lack of strong colocalization support may be caused by an inadequate power, even though the current colocalization analysis was based on a large-scale meta-analysis GWAS including 107,167 type 2 diabetes cases. In addition, this study did not pinpoint certain proteins associated with diabetes described in previous studies, like growth-differentiation factor-15 and matrix metallopeptidase 12,13 which did not indicate that our findings were against these associations. Instead, these associations were also observed in our analysis, albeit non-significantly, which might be caused by a heavy multiple testing burden. However, this multiple-testing correction strategy suits one of the study’s aims, which is to discover proteins strongly associated with type 2 diabetes.

Except for the GCKR protein, a well-known protein affecting diabetes risk,5 our study revealed many other plasma proteins that might be causally associated with the risk of type 2 diabetes based on a larger pool of plasma proteins. For example, HHIP has been found to be involved in the development of cancers.19 Our study in line with an observational study20 established a positive association between HHIP protein levels and the risk of type 2 diabetes, which indicates that the Hedgehog signaling pathway may play a role in glycemic homeostasis. HIBCH plays a role in the catabolism of the branched-chain amino acid valine that has been positively associated with the risk of type 2 diabetes.21 In addition, this protein also determines the metabolism of 3-hydroxyisobutyrate, which was associated with incident type 2 diabetes.22 Altogether, support our finding on a causal

<table>
<thead>
<tr>
<th>Study</th>
<th>Phenotype</th>
<th>Cases</th>
<th>Controls</th>
<th>PMID</th>
<th>Adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td>UKB-PPP</td>
<td>Plasma protein</td>
<td>54,306</td>
<td>–</td>
<td>Preprint</td>
<td>age, gender, batch, UKB center, UKB genetic array, time between blood sampling and measurement and the first 20 genetic principal components</td>
</tr>
<tr>
<td>deCODE</td>
<td>Plasma protein</td>
<td>35,559</td>
<td>–</td>
<td>34857953</td>
<td>age, gender, and sample age</td>
</tr>
<tr>
<td>DIAGRAM</td>
<td>Type 2 diabetes</td>
<td>74,124</td>
<td>824,006</td>
<td>30297969</td>
<td>principal components, relatedness, and study-specific covariates</td>
</tr>
<tr>
<td>FinnGen R8</td>
<td>Type 2 diabetes</td>
<td>33,043</td>
<td>284,971</td>
<td>NA</td>
<td>age, gender, and ≤ 20 genetic principal components</td>
</tr>
<tr>
<td>GIANET</td>
<td>BMI</td>
<td>681,275</td>
<td>–</td>
<td>30124842</td>
<td>age, gender, recruitment center, genotyping batches, and 10 PCs</td>
</tr>
<tr>
<td>Neale Lab</td>
<td>BMI</td>
<td>336,107</td>
<td>–</td>
<td>NA</td>
<td>age, gender, 10 genetic principal components</td>
</tr>
<tr>
<td>FinnGen R8</td>
<td>Diabetic ketoacidosis</td>
<td>7,201</td>
<td>249,480</td>
<td>NA</td>
<td>age, gender, and ≤ 20 genetic principal components</td>
</tr>
<tr>
<td>FinnGen R8</td>
<td>Diabetic retinopathy</td>
<td>8,942</td>
<td>283,545</td>
<td>NA</td>
<td>age, gender, and ≤ 20 genetic principal components</td>
</tr>
<tr>
<td>FinnGen R8</td>
<td>Diabetic nephropathy</td>
<td>3,676</td>
<td>283,456</td>
<td>NA</td>
<td>age, gender, and ≤ 20 genetic principal components</td>
</tr>
<tr>
<td>FinnGen R8</td>
<td>Diabetic neuropathy</td>
<td>2,444</td>
<td>249,480</td>
<td>NA</td>
<td>age, gender, and ≤ 20 genetic principal components</td>
</tr>
<tr>
<td>FinnGen R8</td>
<td>Diabetic maculopathy</td>
<td>3,115</td>
<td>283,472</td>
<td>NA</td>
<td>age, gender, and ≤ 20 genetic principal components</td>
</tr>
<tr>
<td>FinnGen R8</td>
<td>Diabetic hypoglycemia</td>
<td>6,500</td>
<td>249,480</td>
<td>NA</td>
<td>age, gender, and ≤ 20 genetic principal components</td>
</tr>
<tr>
<td>FinnGen R8</td>
<td>Diabetes peripheral circulatory complications</td>
<td>1,975</td>
<td>283,224</td>
<td>NA</td>
<td>age, gender, and ≤ 20 genetic principal components</td>
</tr>
<tr>
<td>FinnGen R8</td>
<td>Cardiomyopathy</td>
<td>5,344</td>
<td>237,263</td>
<td>NA</td>
<td>age, gender, and ≤ 20 genetic principal components</td>
</tr>
<tr>
<td>CARDioGRAMplusC4D</td>
<td>Coronary artery disease</td>
<td>60,801</td>
<td>123,504</td>
<td>26343387</td>
<td>age and gender</td>
</tr>
<tr>
<td>Malik et al. GWAS15</td>
<td>Ischemic stroke</td>
<td>34,217</td>
<td>406,111</td>
<td>29531354</td>
<td>age and gender</td>
</tr>
</tbody>
</table>

CARDioGRAMplusC4D, Coronary Artery Disease Genome wide Replication and Meta-analysis (CARDioGRAM) plus The Coronary Artery Disease (C4D) Genetics; GIANET, The Genetic Investigation of ANthropometric Traits consortium; NA, not available; UKB-PPP, The UK Biobank Pharma Proteomics Project.
inverse association between HIBCH protein and type 2 diabetes. Likewise, MAPK8-related signaling was found to be related to the development of type 2 diabetes in animal studies. The causality of this association in human has now been strengthened by our study. Another interesting protein associated with diabetes was MPPED2, which was found to impact the development of cancers with functions of regulating cell proliferation, differentiation, and apoptosis. However, whether the MPPED2-diabetes association was caused by pleiotropy needs assessment. PGM1 protein has been proven to act in glucose metabolism and its encoding gene has been identified as a shared loci between type 1 and type 2 diabetes. Thus, the association between PGM1 and type 2 diabetes can be regarded as a positive control. More studies are needed to verify our findings.

In reverse MR analysis, we observed associations of genetic liability to type 2 diabetes with six proteins, including some well-established diabetes-associated proteins (i.e., SHBG, MICB, and HIBCH), which raises the concern that these proteins may be associated with type 2 diabetes due to reverse causation. However, we noticed that the directions of protein-diabetes and diabetes-protein associations were opposite, except that for chymotrypsinogen B2, which minimizes the possibility of reverse causation and implies potential bidirectional associations caused by a feedback mechanism.

Our analysis of diabetes-associated proteins in relation to diabetic complications added insights on diabetic progression from...
We found that diabetes-associated proteins seemed to be more associated with microvascular complications compared to macrovascular complications, which supported the findings that the associations of diabetes with macrovascular outcomes were less pronounced, and significant than those with microvascular. However, given that data used on three major macrovascular complications were obtained from general populations instead of diabetic patients, our result might be influenced by this.

We found that four proteins (i.e., HLA-DRA, AGER, HSPA1A, and HSPA1B) were associated with most studied diabetic complications. However, it is important to note that these protein associations with diabetic complications lacked strong colocalization support, possibly due to limited statistical power. HLA-DR consisted of HLA-DRA and HLA-DRB1 (HLA class II histocompatibility antigen, DRB1 beta chain) has been associated with the risk of type 1 diabetes. A few studies linked this protein to the development of type 2 diabetes as well as its progression. A study in the Han ethnicity of China showed that alleles of HLA-DRA and HLA-DRB1 genes were associated with the susceptibility of type 2 diabetes and diabetic nephropathy. These associations were replicated in a dataset-based multi-omics study. Another study found that HLA-DR-related inflammation might be associated with beta-cell destruction in patients with type 2 diabetes, which may explain the associations with diabetic complications. Still, the immune response to increased levels of HLA-DRA is complex. Our study can only convey that HLA-DRA-related immune reactions may play a role in diabetic progression and more studies are needed to explore in detail the underlying mechanistic pathways. AGER is a cell-surface...
receptor of AGEs (advanced glycation end products) by glycation. It will activate pro-inflammatory pathways when AGEs bind to AGER.32 Thus, low levels of AGER may decrease glycation-related inflammation and thus decrease the risk of diabetes and its complications, which is supported by our findings. Many small molecule AGER inhibitors or antagonists have been proposed to treat diabetic complications and cancer progression with a good tolerance in humans.33 Heat shock proteins were associated with hypoglycemia34 and diabetic foot ulcer.35 However, the mechanisms underlying the inverse associations of HSPA1A and HSPA1B with diabetes and its complications are unclear, but possibly related to attenuation of oxidative stress.36

Except for the above four proteins, we also observed GCKR and ANGPTL4 associated with some complications, which was consistent with recent trial evidence that GCKR and ANGPTL4 are currently in therapeutic development for metabolic disorders. Our data demonstrated potential involvement of these two proteins in the development of type 2 diabetes and related complications; thus, these drugs under development hold also promise for these indications.

This study has several strengths, including a large pool of plasma proteins, a large number of type 2 diabetes cases, a mutual validation in two independent outcome data, and a supportive analysis of colocalization. Limitations deserve to be discussed when interpreting our results. First, this analysis was confined to Europeans, which limited the generalizability of our findings to other populations. Second, we did colocalization analysis only for proteins included in the deCODE Health study37 due to data availability. Thus, whether the MR associations for proteins with instruments from the UKB-PPP study are supported by colocalization analysis remain unknown. Third, even though we meta-analyzed GWAS data on type 2 diabetes from two independent sources to increase the power, we may have overlooked some weak associations, especially in colocalization analyses. Fourth, the power may have been inadequate in the analysis of diabetic complications due to a small number of cases. Fifth, data on coronary artery disease, ischemic stroke, and cardiomyopathy were not obtained from a diabetic population, which might influence the comparison between the associations for macrovascular and microvascular complications with diabetes-associated proteins. Sixth, even though we were not able to differentiate the diabetic complications of type 1 or type 2 diabetes, given that the analysis was based on the summary-level data, our results should reflect more on the complications of type 2 diabetes due to a small number of complication cases of type 1 in the FinnGen study. For example, there were 1,897 cases of type 1 diabetes with ketoacidosis among 7,201 individuals with ketoacidosis. Seventh, even though cis-MR analysis can minimize the influence of horizontal pleiotropy, the observed associations might be affected by this phenomenon, especially for a certain protein with an encoding gene exerting a broad range of effects. Eighth, our study could not assess the risk of non-specific binding for proteins or the variability of protein measurement since this study was based on summary-level data. However, including proteins with cis-SNPs might decrease the risk of measurement error due to non-specific binding and results were overall consistent for overlapping proteins with data measured by SomaScan and Olink. Last, but not least, we used proteomic data from Icelanders and outcome data from Finnish, which might introduce bias by violating the second assumption of MR because the genetic backgrounds of these two populations may differ from other European populations. However, this potential bias is likely to be minimal as we observed directionally consistent protein-type 2 diabetes associations with outcome data from DIAGRAM and FinnGen studies. Furthermore, we observed that more than 85% of the proteins-type 2 diabetes associations with a p < 0.05 were directionally consistent among proteins with genetic instruments from both deCODE (Icelanders) and UKB-PPP (British). The around 15% inconsistency might be attributed to different genetic architectures but more likely to different proteomic profiling platforms (SomaScan and Olink) used in the two cohorts. This hypothesis is supported by studies that observed that at least 14.7% of proteins had poor correlations when measured in SomaScan and Olink platforms.38–40

In summary, this proteome-wide MR and colocalization analysis identified many plasma proteins highly causally associated with type 2 diabetes. Among these proteins, almost one-third were associated with diabetic complications, in particular acute and microvascular complications. Universal effects of four diabetes-associated proteins (HLA-DRA, AGER, HSPA1A, and HSPA1B) on most studied diabetic complications may suggest the roles of inflammation and oxidative stress in diabetic progression. Our findings also suggest that AGER inhibitors or antagonists may be a promising therapeutic target for diabetic complications.

Limitations of the study
Our findings, in relation to the analysis itself, may be subject to the following potential limitations: (1) bias due to population structure; (2) macrovascular complications in non-diabetic populations; (3) the presence of complications from both type 1 and type 2 diabetes; (4) horizontal pleiotropy; (5) heterogeneity introduced by the usage of two protein profiling platforms; and (6) limited generalizability. Each of these factors has been thoroughly addressed in the text. Furthermore, as these associations were established through in silico analyses, they require further validation through animal studies and population-based research. Despite the inclusion of a vast number of blood proteins in this study, we may have inadvertently overlooked significant proteins that lack genetic instruments. The risk prediction ability of these identified proteins also warrants further studies.

STAR METHODS
Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODELS AND SUBJECT DETAILS
- METHOD DETAILS
  - Data sources for plasma proteins
  - Data sources for type 2 diabetes
SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2023.101174.

ACKNOWLEDGMENTS

The current study received no funding. We acknowledge investigators and participants in the UK Biobank study, the deCODE project, the FinnGen study, and cited genome-wide association studies for sharing data.

AUTHOR CONTRIBUTIONS

S.Y., F.X., C.S.M., and S.C.L. conceived and designed the study. S.Y. and F.X. undertook the statistical analyses. S.Y. wrote the first draft of the manuscript. S.Y., F.X., X.L., J.C., J.Z., C.S.M., and S.C.L. interpreted the data, provided important comments to the manuscript, and approved the final version of the manuscript.

DECLARATION OF INTERESTS

All authors declare no conflict of interest.

REFERENCES

**STARMETHODS**

**KEY RESOURCES TABLE**

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deposited data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasma protein</td>
<td>UKB-PPP</td>
<td><a href="https://www.ukbiobank.ac.uk/">https://www.ukbiobank.ac.uk/</a></td>
</tr>
<tr>
<td>Plasma protein</td>
<td>deCODE</td>
<td><a href="https://www.decode.com/summarydata/">https://www.decode.com/summarydata/</a></td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>DIAGRAM</td>
<td><a href="https://www.diagram-consortium.org/">https://www.diagram-consortium.org/</a></td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>Finngen R8</td>
<td><a href="https://www.finngen.fi/fi">https://www.finngen.fi/fi</a></td>
</tr>
<tr>
<td>Body mass index</td>
<td>GIANT</td>
<td><a href="https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files">https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files</a></td>
</tr>
<tr>
<td>Body mass index</td>
<td>Neale Lab</td>
<td><a href="http://www.nealelab.is/uk-biobank">http://www.nealelab.is/uk-biobank</a></td>
</tr>
<tr>
<td>Diabetic ketoacidosis</td>
<td>Finngen R8</td>
<td><a href="https://www.finngen.fi/fi">https://www.finngen.fi/fi</a></td>
</tr>
<tr>
<td>Diabetic retinopathy</td>
<td>Finngen R8</td>
<td><a href="https://www.finngen.fi/fi">https://www.finngen.fi/fi</a></td>
</tr>
<tr>
<td>Diabetic nephropathy</td>
<td>Finngen R8</td>
<td><a href="https://www.finngen.fi/fi">https://www.finngen.fi/fi</a></td>
</tr>
<tr>
<td>Diabetic neuropathy</td>
<td>Finngen R8</td>
<td><a href="https://www.finngen.fi/fi">https://www.finngen.fi/fi</a></td>
</tr>
<tr>
<td>Diabetic maculopathy</td>
<td>Finngen R8</td>
<td><a href="https://www.finngen.fi/fi">https://www.finngen.fi/fi</a></td>
</tr>
<tr>
<td>Diabetic hypoglycemia</td>
<td>Finngen R8</td>
<td><a href="https://www.finngen.fi/fi">https://www.finngen.fi/fi</a></td>
</tr>
<tr>
<td>Diabetes peripheral circulatory complications</td>
<td>Finngen R8</td>
<td><a href="https://www.finngen.fi/fi">https://www.finngen.fi/fi</a></td>
</tr>
<tr>
<td>Cardiomyopathy</td>
<td>Finngen R8</td>
<td><a href="https://www.finngen.fi/fi">https://www.finngen.fi/fi</a></td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>CARDIoGRAMplusC4D</td>
<td><a href="http://www.cardiogramplusc4d.org/">http://www.cardiogramplusc4d.org/</a></td>
</tr>
<tr>
<td>Ischemic stroke</td>
<td>Malik et al. [15]</td>
<td><a href="https://www.ebi.ac.uk/gwas/publications/29531354">https://www.ebi.ac.uk/gwas/publications/29531354</a></td>
</tr>
<tr>
<td>1000 Genomes European reference</td>
<td>1000 Genomes</td>
<td><a href="https://www.internationalgenome.org/data/">https://www.internationalgenome.org/data/</a></td>
</tr>
</tbody>
</table>

**SOFTWARE AND ALGORITHMS**

<table>
<thead>
<tr>
<th>Software and algorithms</th>
<th>Github</th>
<th><a href="https://mrcieu.github.io/TwoSampleMR/">https://mrcieu.github.io/TwoSampleMR/</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>MendelianRandomization</td>
<td>R project</td>
<td><a href="https://cran.r-project.org/web/packages/MendelianRandomization/index.html">https://cran.r-project.org/web/packages/MendelianRandomization/index.html</a></td>
</tr>
<tr>
<td>Coloc R package</td>
<td>Github</td>
<td><a href="https://github.com/chriswallace/coloc">https://github.com/chriswallace/coloc</a></td>
</tr>
<tr>
<td>SuSiE R package</td>
<td>Github</td>
<td><a href="https://stephenslab.github.io/susieR/">https://stephenslab.github.io/susieR/</a></td>
</tr>
<tr>
<td>GeneMANIA</td>
<td>GeneMANIA</td>
<td><a href="https://genemania.org/">https://genemania.org/</a></td>
</tr>
</tbody>
</table>

**RESOURCE AVAILABILITY**

**Lead contact**
Further information and requests should be directed to the lead contact, Shuai Yuan (shuai.yuan@ki.se).

**Materials availability**
This study did not generate new unique reagents.

**Data and code availability**
This paper analyzes existing, publicly available data. All GWAS are referenced. This paper does not report the original code. Any additional information required to reanalyze the data reported in this report is available from the lead contact upon request.

**EXPERIMENTAL MODELS AND SUBJECT DETAILS**

The analysis was based on summary-level GWAS data in European populations. We first examined the associations of genetically predicted plasma proteins with risk of type 2 diabetes in two independent large-scale studies. To explore whether the diabetes-associated proteins play a role in the development of diabetic complications, we estimated the associations of diabetes-associated proteins with ten diabetic complications. We also tested the associations of the diabetes-associated proteins with BMI and conducted a two-stage network MR analysis to reveal whether these protein-diabetes associations were mediated via obesity. The used studies had been approved by corresponding ethical review committees and participants had signed the consent form. This study was approved by the Swedish Ethical Review Authority.
METHOD DETAILS

Data sources for plasma proteins
We selected index cis-SNPs associated with the levels of plasma proteins at the genome-wide significance level (p < 5 × 10^{-8}) as instrumental variables from two large-scale GWASs in the UK Biobank Pharma Proteomics Project (UKB-PPP) and the deCODE Health study. Cis-SNPs were defined as SNPs within 1Mb from the gene encoding the protein, and linkage disequilibrium was estimated based on 1000 Genomes European panel. UKB-PPP conducted proteomic profiling on blood plasma samples from 54,306 participants using the Olink platform and collected data on 1,463 proteins. For the two-sample MR analysis, we selected index cis-SNPs as instrumental variables for 1161 proteins. Likewise, index cis-SNPs for 1423 plasma proteins were obtained from the deCODE Health study where 4,907 aptamers were measured among 35,559 Icelanders using the SomaScan platform. There were 509 proteins with index cis-SNPs that overlapped between the two studies. Given that colocalization analysis could only be performed for proteins from the deCODE Health study due to the availability of the summary-level data, we used genetic instrumental variables selected from the deCODE Health study for these overlapped proteins in the main results. We have presented results for the overlapping proteins with genetic instruments from two studies to examine the consistency of results based on different proteomic profiling platforms.

Data sources for type 2 diabetes
Data on the associations of protein-associated SNPs with type 2 diabetes were obtained from the DIAGRAM (DIAbetes Genetics Replication And Meta-analysis) consortium and the FinnGen study. The DIAGRAM consortium included 32 studies with a total of 74,124 cases and 824,006 controls of European descent. We used the latest release data on type 2 diabetes from the FinnGen study R8 in this analysis, which comprised 33,043 cases 284,971 controls. There were no sample overlaps between two outcome datasets. In MR analysis, we treated the DIGRAM consortium as the discovery study and the FinnGen R8 study as the replication. To increase the power, we meta-analyzed the two GWASs and performed colocalization analysis based on the GWAS meta-analysis data (107,167 cases and 1,108,976 controls). The GWAS meta-analysis was performed by the METAL software. Genetic variants associated with type 2 diabetes at p < 5 × 10^{-8} in this GWAS meta-analysis and with low linkage disequilibrium (R^2 < 0.001) were selected as the instrument variable for type 2 diabetes in the reverse MR analysis.

Data sources for BMI
Obesity is an important causal risk factor for type 2 diabetes. We estimated the associations of diabetes-associated proteins with BMI to reveal whether these associations may be mediated via obesity. Summary-level data on BMI were available from a GWAS in the GIANT (Genetic Investigation of ANthropometric Traits) consortium that included 681,275 individuals. Given a comparatively small number of SNPs analyzed in this genome-wide association analysis (causing missing in MR analysis), we also used GWAS data on BMI (N = 336,107) from the Neale Lab. The Neale Lab GWAS on BMI was based on 336,107 participants from the UK Biobank study. For proteins associated with BMI and type 2 diabetes in a consistent direction, we performed a two-stage network MR analysis using the Neale Lab data and GWAS meta-analysis of type 2 diabetes to estimate the mediation of BMI in the protein-diabetes associations. Genetic variants associated with BMI at p < 5 × 10^{-8} in Neale Lab data and with low linkage disequilibrium (R^2 < 0.001) were selected as the instrument variable for BMI in this network MR analysis.

Detailed methods of the two-stage network MR analysis and mediation calculation can be found elsewhere.

Data sources for diabetic complications
The analysis included ten diabetic complications, which can be classified into four categories that are acute complications (diabetic ketoacidosis), diabetic hypoglycemia, chronic microvascular complications (diabetic retinopathy, diabetic nephropathy, diabetic neuropathy, and diabetic maculopathy), and chronic macrovascular complications (diabetes peripheral circulatory complications, cardiomyopathy, coronary artery disease, and ischemic stroke). Given no available data on coronary artery disease and ischemic stroke among diabetic patients, we used data on coronary artery disease (60,801 cases and 123,504 controls) and ischemic stroke (34,217 cases and 406,111 controls) from corresponding genome-wide association studies among general populations, respectively. Likewise, data on general cardiomyopathy were obtained from the FinnGen R8 study (5344 cases). Summary-level data on other diabetic complications were obtained from the FinnGen R8 study with the number of cases ranging from 1975 for diabetic neuropathy, and diabetic maculopathy, and chronic macrovascular complications (diabetes peripheral circulatory complications, cardiomyopathy, coronary artery disease, and ischemic stroke). Given no available data on coronary artery disease and ischemic stroke among diabetic patients, we used data on coronary artery disease (60,801 cases and 123,504 controls) and ischemic stroke (34,217 cases and 406,111 controls) from corresponding genome-wide association studies among general populations, respectively. Likewise, data on general cardiomyopathy were obtained from the FinnGen R8 study (5344 cases). Summary-level data on other diabetic complications were obtained from the FinnGen R8 study with the number of cases ranging from 1975 for diabetic neuropathy, and diabetic maculopathy, and chronic macrovascular complications (diabetes peripheral circulatory complications, cardiomyopathy, coronary artery disease, and ischemic stroke). Given no available data on coronary artery disease and ischemic stroke among diabetic patients, we used data on coronary artery disease (60,801 cases and 123,504 controls) and ischemic stroke (34,217 cases and 406,111 controls) from corresponding genome-wide association studies among general populations, respectively. Likewise, data on general cardiomyopathy were obtained from the FinnGen R8 study (5344 cases). Summary-level data on other diabetic complications were obtained from the FinnGen R8 study with the number of cases ranging from 1975 for peripheral circulatory complications to 8942 for diabetic retinopathy.

QUANTIFICATION AND STATISTICAL ANALYSIS

MR analysis
Protein-associated SNPs unavailable in outcome data were replaced by SNP proxies with a high linkage disequilibrium (R^2 ≥ 0.8) based on 1000 Genomes European reference panel. Missing SNPs without suitable SNP proxies were removed from the analysis (N = 190). The F statistic was calculated to assess the strength of the instrumental variables. We estimated statistical power for the analyses assuming a disease prevalence of 2% in general. The ORs and corresponding CIs of the associations between plasma proteins and studied outcomes were estimated by the Wald ratio and the delta method (a general method for deriving the variance of
a function of asymptotically normal random variables with known variance and commonly used to estimate standard error of MR ratio estimate), \(^4\) respectively. The associations for one protein from two outcome data were meta-analyzed using the fixed-effect model. The MR associations were scaled to one SD increase in genetically predicated levels of circulating proteins. Bonferroni method was used to correct for multiple testing in the analysis for type 2 diabetes. The combined association with a \(p\) value \(<2.65 \times 10^{-5}\) (0.05/1885 proteins) was deemed significant. For type 2 diabetes, we performed a reverse MR analysis to estimate the associations of type 2 diabetes liability with the levels of identified proteins to explore potential reverse causation. For the MR analysis of diabetic complications, we used the false discovery rate approach to correct for multiple testing given a comparative small sample size. MR analyses were performed using TwoSampleMR and MendelianRandomization packages in R software (4.4.1).

**Colocalization analysis**

We performed the colocalization analysis to test whether identified associations of proteins with type 2 diabetes and diabetic complications were driven by linkage disequilibrium. The analysis was based on a Bayesian model that assesses the support for five exclusive hypotheses: 1) no association with either trait; 2) association with trait 1 only; 3) association with trait 2 only; 4) both traits are associated, but distinct causal variants were for two traits; and 5) both traits are associated, and the same shares causal variant for both traits. \(^5\) A posterior probability is provided for each hypothesis testing (H0, H1, H2, H3, and H4). In this analysis, we used the setting that is the prior probabilities of the SNP being associated with trait 1 only (\(p1\)) at \(1 \times 10^{-6}\); the probability of the SNP being associated with trait 2 only (\(p2\)) at \(1 \times 10^{-4}\); and the probability of the SNP being associated with both traits (\(p12\)) at \(1 \times 10^{-5}\). Two signals were considered to have a strong support of colocalization if the posterior probability for shared causal variants (PH4) was \(\geq 0.8\). Medium colocalization indication was defined as \(0.5 < \text{PH4} < 0.8\). The analysis was performed using the coloc package in R software (4.4.1). Given that the traditional colocalization approach cannot detect the scenario where the exposure and outcome traits share more than one causal hit, we, therefore, utilized SuSiE (Sum of Single Effects) \(^6\) colocalization method by integrating proteomics GWAS summary statistics and genetic correlation matrix reference panels based on individuals of European ancestry from the 1000 Genomes Project Phase 3 to identify multiple causal variants.

**Function and network prediction**

We used GeneMANIA (http://www.genemania.org) to predict the functions and networks of cis-genes for the diabetes-associated proteins and both diabetes- and BMI-associated proteins. Detailed information on included datasets in GeneMANIA is described somewhere else. \(^7\) Enriched function pathways were considered to be significant with false discovery rate \(<0.05\).