Diabetologia (2023) 66:1655–1668
https://doi.org/10.1007/s00125-023-05943-2

ARTICLE

Associations of plasma proteomics with type 2 diabetes and related traits: results from the longitudinal KORA S4/F4/FF4 Study

Hong Luo[1,2] · Alina Bauer[1] · Jana Nano[1,2] · Agnese Petrera[3] · Wolfgang Rathmann[4,5] · Christian Herder[5,6,7] · Stefanie M. Hauck[3,8] · Benjamin B. Sun[9] · Annika Hoyer[10] · Annette Peters[1,2,8] · Barbara Thorand[1,8]

Received: 19 October 2022 / Accepted: 12 April 2023 / Published online: 13 June 2023

Abstract

Aims/hypothesis This study aimed to elucidate the aetiological role of plasma proteins in glucose metabolism and type 2 diabetes development.

Methods We measured 233 proteins at baseline in 1653 participants from the Cooperative Health Research in the Region of Augsburg (KORA) S4 cohort study (median follow-up time: 13.5 years). We used logistic regression in the cross-sectional analysis (n=1300), and Cox regression accounting for interval-censored data in the longitudinal analysis (n=1143). We further applied two-level growth models to investigate associations with repeatedly measured traits (fasting glucose, 2 h glucose, fasting insulin, HOMA-B, HOMA-IR, HbA_{1c}), and two-sample Mendelian randomisation analysis to investigate causal associations. Moreover, we built prediction models using priority-Lasso on top of Framingham-Offspring Risk Score components and evaluated the prediction accuracy through AUC.

Results We identified 14, 24 and four proteins associated with prevalent prediabetes (i.e. impaired glucose tolerance and/or impaired fasting glucose), prevalent newly diagnosed type 2 diabetes and incident type 2 diabetes, respectively (28 overlapping proteins). Of these, IL-17D, IL-18 receptor 1, carbonic anhydrase-5A, IL-1 receptor type 2 (IL-1RT2) and matrix extracellular phosphoglycoprotein were novel candidates. IGF binding protein 2 (IGFBP2), lipoprotein lipase (LPL) and paraoxonase 3 (PON3) were inversely associated while fibroblast growth factor 21 was positively associated with incident type 2 diabetes. LPL was longitudinally linked with change in glucose-related traits, while IGFBP2 and PON3 were linked with changes in both insulin- and glucose-related traits. Mendelian randomisation analysis suggested causal effects of LPL on type 2 diabetes and fasting insulin. The simultaneous addition of 12 priority-Lasso-selected biomarkers (IGFBP2, IL-18, IL-17D, complement component C1q receptor, V-set and immunoglobulin domain-containing protein 2, IL-1RT2, LPL, CUB domain-containing protein 1, vascular endothelial growth factor D, PON3, C-C motif chemokine 4 and tartrate-resistant acid phosphatase type 5) significantly improved the predictive performance (ΔAUC 0.0219; 95% CI 0.0052, 0.0624).

Conclusions/interpretation We identified new candidates involved in the development of derangements in glucose metabolism and type 2 diabetes and confirmed previously reported proteins. Our findings underscore the importance of proteins in the pathogenesis of type 2 diabetes and the identified putative proteins can function as potential pharmacological targets for diabetes treatment and prevention.

Keywords Cohort study · Mendelian randomisation · Proteomics · Traits of glucose and insulin · Type 2 diabetes

Abbreviations

CA5A	Carbonic anhydrase-5A
FDR	False discovery rate
FGF21	Fibroblast growth factor 21
FORS	Framingham-Offspring Risk Score
GWAS	Genome-wide association study
IGFBP1	IGF binding protein 1
IGFBP2	IGF binding protein 2
IL-18R1	IL-18 receptor 1
IL-1RT2	IL-1 receptor type 2
IV	Instrumental variable
KORA	Cooperative Health Research in the Region of Augsburg
LOD	Limit of detection

✉ Barbara Thorand
barbara.thorand@helmholtz-munich.de

Extended author information available on the last page of the article

Research in context

What is already known about this subject?

- Several studies have investigated the link between proteins and type 2 diabetes using proteomic approaches that simultaneously measure a large number of proteins, opening new avenues in biomarker discovery
- Few studies investigated the longitudinal association between proteins and diabetes-related traits
- There is little consensus on a specific biomarker set that would be most relevant to improving incident type 2 diabetes prediction

What is the key question?

- Using a high-throughput proteomic technology, can we discover new proteins associated with type 2 diabetes and related traits and, if so, are the identified proteins causally related to type 2 diabetes development?

What are the new findings?

- A total of 28 overlapping proteins, of which five were novel, were associated with prevalent prediabetes (14 proteins), prevalent newly diagnosed type 2 diabetes (24 proteins) or incident type 2 diabetes (four proteins). Three incident type 2 diabetes-related proteins were linked to the longitudinal change in glucose- and/or insulin-related traits
- Mendelian randomisation analysis suggested causal effects of lipoprotein lipase on type 2 diabetes and fasting insulin
- The combination of IGFBP2, IL-18, IL-17D, CD93, VSIG2, IL-1RT2, LPL, CDCP1, VEGFD, PON3, CCL4 and TR-AP improved diabetes prediction on top of classical risk factors

How might this impact on clinical practice in the foreseeable future?

- Our epidemiological findings further elucidated the role of proteins in the pathogenesis of type 2 diabetes and discovered new potential drug targets

LPL	Lipoprotein lipase
MEPE	Matrix extracellular phosphoglycoprotein
MR	Mendelian randomisation
NPX	Normalised protein expression
PEA	Proximity extension assay
*p*_FDR	*p* value after controlling for FDR
PON3	Paraoxonase 3
PPI	Protein–protein interaction

Introduction

Type 2 diabetes burden remains a major public health concern with a considerable impact on quality of life and health expenditures. Early diabetes screening and lifestyle interventions provide an opportunity to halt or delay disease onset [1]. Despite intense research on the pathophysiology of diabetes, the underlying mechanisms are not fully elucidated. Identification of novel biomarkers linked to the development of type 2 diabetes and early derangements in glucose metabolism may offer the opportunity to further advance our knowledge not only in uncovering aetiology, but also in improving disease prevention and prediction.

Proteins are the crucial functional units in biological processes. Most previous studies linked single proteins to type 2 diabetes, limiting the possibility to identify novel interconnected pathways. Advances in proteomic technology make it possible to simultaneously measure a large number of proteins, opening new avenues in biomarker discovery. Previous epidemiological studies using high-throughput proteomic technologies have identified up to 142 plasma proteins associated with prevalent type 2 diabetes [2–6], and additional ones for prevalent prediabetes (i.e. impaired glucose tolerance and/or impaired fasting glucose) [7, 8] or insulin resistance [9, 10]. However, most of the studies until now were of a cross-sectional nature, without the opportunity to elucidate temporality. Of note, the limited number of longitudinal proteomics studies [2, 5, 9, 11–13] generally identified only a few proteins to be statistically significantly associated with incident type 2 diabetes. Given the dynamic nature of glucose and insulin metabolism prior to diabetes development, it is also important to investigate the role of

proteomics in longitudinal changes in glucose and related traits. Furthermore, although several studies have shown that the addition of newly identified biomarkers improves prediction models of type 2 diabetes [5, 9, 11, 12, 14, 15], most studies lacked replication and there is still little consensus on a specific biomarker set that would be most relevant to improving type 2 diabetes prediction.

Therefore, using the proximity extension assay (PEA) technology, we aimed to investigate the association of 233 plasma proteins with prevalent prediabetes as well as newly diagnosed type 2 diabetes and with type 2 diabetes development and six closely related traits (fasting glucose, fasting insulin, 2 h glucose, HOMA-IR, HOMA-B and HbA_{1c}), in a cohort with up to three repeated measurements of the examined traits. Moreover, we performed Mendelian randomisation (MR) analysis to investigate the directionality of the observed associations. Lastly, we investigated type 2 diabetes prediction performance of the identified biomarkers on top of the Framingham-Offspring Risk Score (FORS) variables [16].

Methods

Study population

We used data from the population-based Cooperative Health Research in the Region of Augsburg (KORA) S4 survey (1999–2001) (n=4261), F4 (2006–2008) (n=3080) and FF4 (2013/2014) (n=2279) [17]. The present study population was restricted to participants aged 55–74 in the S4 study (n=1653) due to data availability. Participants with known diabetes or unclear diabetes diagnoses were excluded from the cross-sectional analysis. Furthermore, we excluded those who were non-fasting and had missing data as shown in electronic supplementary material (ESM) Fig. 1. Thus, the cross-sectional analysis finally included 1300 participants. For the prospective analysis, we additionally excluded those who were newly diagnosed with type 2 diabetes based on OGTT at baseline (S4). After further exclusion of 41 participants with unclear information on the diagnosis of diabetes during follow-up, 1143 participants remained for prospective analyses regarding incident type 2 diabetes (n=178). Of these, 881 participants attended the KORA F4 and/or FF4 follow-up examinations and had complete information to ascertain diabetes status. In addition to the follow-up examinations F4 and FF4, written questionnaires were distributed to all participants to assess their diabetes status, date of diagnosis and whether the disease had been diagnosed by a physician in 2008/2009 and 2016 [17]. Information from these questionnaires was used for another 262 participants to assess diabetes status. The longitudinal analyses of the traits of blood glucose and insulin were restricted to 840–896 participants with baseline data and at least one additional measurement at F4 or FF4 of the respective trait (see ESM Fig. 1).

Proteomics measurements

Protein levels were measured in plasma samples from KORA S4 using the PEA technology developed by Olink (Olink Proteomics, Uppsala, Sweden). Three panels (CVD-II, CVD-III and Inflammation), each comprising 92 protein biomarkers, were measured as described previously [18]. The panels of biomarkers were selected due to the potential importance of CVD and inflammation for type 2 diabetes pathophysiology based on prior knowledge from experimental and epidemiological studies. Briefly, the Olink platform provided $\log_2$-normalised protein expression (NPX) values and these were divided by their respective SDs, calculated in the complete dataset prior to exclusions. We excluded 29 biomarkers with values below the limit of detection (LOD) in >25% of all participants (all remaining values <LOD were retained in the data and were not substituted), nine biomarkers duplicated in two panels (four of CVD-II, three of CVD-III and two of Inflammation; the duplicate with more values below the LOD value and a higher inter-assay coefficient of variation was excluded) and five biomarkers that had missing values. Finally, 233 proteins were included in the present analysis.

Outcomes

A detailed description of the measurement methods of outcomes and covariates can be found in the ESM Methods.

All participants without known diabetes received a standard 75 g OGTT after an overnight fast of at least 8 h [11]. To avoid the influence of glucose-lowering drug intake and long-term hyperglycaemia, participants with known diabetes were excluded from the cross-sectional analysis. Therefore, unlike in other studies, prevalent diabetes comprises only newly diagnosed diabetes by OGTT test in S4. Prevalent prediabetes and newly diagnosed type 2 diabetes in S4 were defined according to the 1999/2006 WHO criteria (see the ESM Methods). Incident type 2 diabetes was defined by a validated clinical diagnosis of type 2 diabetes initially assessed through self-report at F4, FF4 or questionnaire responses during the follow-up period in participants without prevalent diabetes at baseline. In addition, all participants fulfilling the criteria for newly diagnosed diabetes described above at either F4 or FF4 were considered to have incident type 2 diabetes. At baseline and during follow-up, self-reported information regarding a medical diagnosis of diabetes and the date of diagnosis was validated by contacting the treating physician or medical chart review, and only those without confirmed diabetes received an OGTT [19].

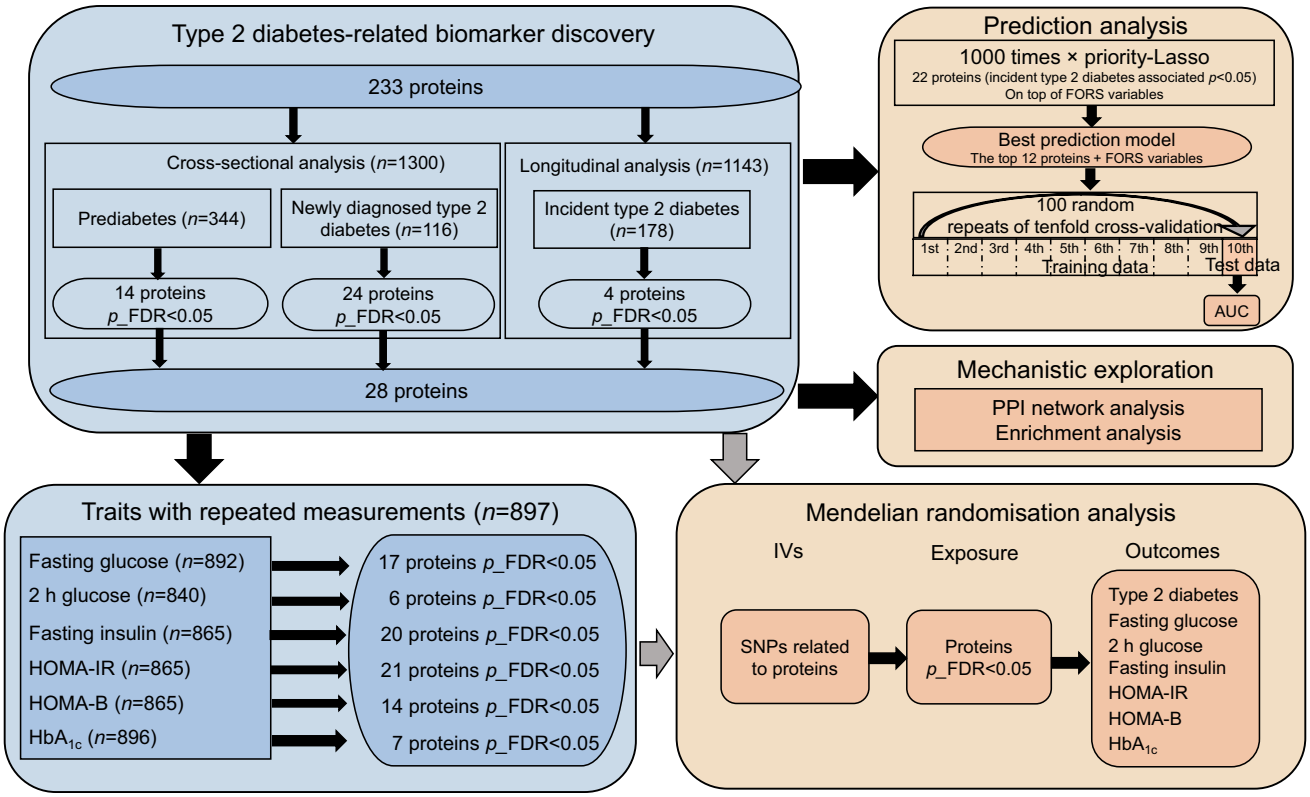


Fig. 1 Flow chart illustrating the analysis strategy

The status changes of participants from KORA S4 to F4 and FF4 are presented in ESM Fig. 2.

Covariates

All participants took part in standard physical and medical examinations at KORA S4. Information about age, sex, parental history of diabetes and medical history was assessed during a standardised interview by trained medical staff. Parental history of diabetes was categorised as positive (at least one parent with diabetes), negative (both parents without diabetes) or unknown diabetes status (other). HDL-cholesterol and triglycerides were measured by standard clinical methods. Waist circumference was evaluated at the minimum abdominal girth. Body weight and height were measured in light clothing by trained investigators. BMI was calculated as body weight (kg) divided by the square of height (m).

Statistical analysis

The analysis strategy of the study is shown in Fig. 1.

Baseline characteristics Characteristics of the study population are shown as mean±SD or median (25th and 75th percentiles) for normally or not normally distributed continuous variables, respectively, and as numbers (percentages) for categorical variables.

Proteome-wide analysis Multivariable logistic regression was used to estimate the associations between each protein and prevalent prediabetes and prevalent newly diagnosed type 2 diabetes (vs normoglycaemia). Cox regression accounting for interval-censoring was used to explore the associations with incident type 2 diabetes. The association analyses were adjusted for important baseline diabetes risk factors, i.e. sex and age (model 1), plus parental history of diabetes, systolic blood pressure, BMI, HDL-cholesterol, triglycerides and waist circumference (model 2). These covariates, together with fasting glucose, are components of the FORS clinical prediction model [16]. As fasting glucose is a defining feature of diabetes, we excluded it from the association analyses and only included it in the prediction analysis. We used the false discovery rate (FDR) (Benjamini–Hochberg method) for each outcome to account for multiple testing. An association was considered statistically significant at a p value<0.05 after controlling for the FDR (p_FDR).

The significant proteins for prevalent prediabetes, prevalent newly diagnosed type 2 diabetes and incident type 2

diabetes were defined as diabetes-related protein biomarkers and were included in the longitudinal analysis of traits of blood glucose and insulin. Protein–protein interaction (PPI) network analysis and enrichment analysis (see the ESM Methods) were conducted for these biomarkers. Two-level growth models [20] (see the ESM Methods) were used to investigate the relationship between the protein biomarkers at S4 and the rate of change from S4 to F4 and FF4 for continuous outcomes. The continuous outcomes (fasting glucose, 2 h glucose, fasting insulin, HOMA-B, HOMA-IR, HbA_{1c}) were log-transformed. Extreme outlier values of fasting insulin at baseline, defined as values above the 98th percentile of the distribution of all insulin measurements at any time point (i.e. >324 pmol/l), were excluded. Models were adjusted for the same covariates as described above.

In sensitivity analyses, we adjusted models regarding incident type 2 diabetes for other diabetes-related lifestyle factors including smoking, physical activity, alcohol use, consumption of whole-grain bread and muesli, consumption of meat and consumption of coffee [14]. Furthermore, in order to explore the potential impact of drug use, we further adjusted for use of lipid-lowering medication at baseline on the incident type 2 diabetes–protein associations and the use of glucose-lowering medication during follow-up on the continuous traits–protein associations. In addition, we excluded the 262 participants who did not participate in F4 or FF4 and only had questionnaire-based information regarding the development of incident type 2 diabetes. We further considered death as a competing risk and used the Fine–Gray subdistribution hazard model to estimate the incidence of type 2 diabetes over time in the presence of death risks. To overcome the effect of early derangements in glucose metabolism, we conducted an association analysis among 840 normoglycaemic individuals at baseline.

Two-sample MR analysis We applied a two-sample MR using published large-scale European genome-wide association studies (GWAS) for selecting instrumental variables (IVs). The details regarding the choice of the GWAS database are shown in the ESM Methods and MR processes are presented in ESM Fig. 3. First, we selected IVs associated with proteins at p value $<5\times10^{-8}$ and restricted these to those in *cis* regions. Second, we clumped the SNPs by using the cut-off $r^2=0.01$, which removed SNPs in linkage disequilibrium with the lead SNP. Third, we removed ambiguous palindromic SNPs (SNPs with A/T or G/C alleles). Finally, we extracted the results of these IVs from the outcome's GWAS.

The Wald ratio test was performed when only one IV was available, whereas the inverse variance-weighted method was performed for proteins with at least two IVs [21, 22]. Cochran's Q test and MR-Egger regression were used to test instrument heterogeneity and directional horizontal pleiotropy. The significance p value was defined as 0.05 divided by the number of tested proteins (Bonferroni correction).

Prediction of incident type 2 diabetes We performed priority-Lasso to deal with the multicollinearity of included variables [23]. Priority-Lasso is a least absolute shrinkage and selection operator (Lasso)-based intuitive analysis strategy that constructs a prediction model for a clinical outcome by defining the blocks of different types of predictor variables. In this study, we defined the nine clinical FORS variables as block 1 and forced block 1 in each repeat, while all 22 proteins nominally significantly associated with incident type 2 diabetes in model 2 were defined as block 2. The penalisation parameter λ values were determined as values with maximum AUC estimated in a tenfold cross-validation. The biomarkers were ranked according to the selection times on the priority-Lasso path. The proteins with a selection frequency >20% among 1000 selection rounds were subsequently added consecutively to the FORS model (with nine clinical variables). To quantify the predictive performance of each built model, the AUC of the FORS model (AUC_{basic}), a model additionally including protein markers ($AUC_{extended}$) and ΔAUC ($AUC_{extended}-AUC_{basic}$) were estimated through tenfold cross-validation [24]. To account for the randomness in the selection process and to reduce the chance of overfitting, the whole process was bootstrapped 100 times.

Data analysis was conducted by using R version 4.1 (https://www.r-project.org/).

Results

Description of the study population

The median follow-up time of this study was 13.5 years. Table 1 presents the characteristics of the study participants at baseline. Among the 1300 participants, 344 and 116 participants had prevalent prediabetes and newly diagnosed type 2 diabetes, respectively, whereas 840 participants were normoglycaemic. Detailed information on six outcome traits in KORA S4, F4 and FF4 is shown in ESM Table 1.

Associations with three type 2 diabetes-related outcomes

Fourteen, 24 and four protein biomarkers were statistically significantly associated with prevalent prediabetes, prevalent newly diagnosed type 2 diabetes and incident type 2 diabetes at p_FDR <0.05, respectively (ESM Tables 2,

3). IGF binding protein 2 (IGFBP2), lipoprotein lipase (LPL) and paraoxonase 3 (PON3) were inversely associated while fibroblast growth factor 21 (FGF21) was positively associated with incident type 2 diabetes. The ORs/HRs and the 95% CIs of the identified 28 protein markers and the overlap of statistically significant markers are shown in Fig. 2. The correlation between the identified 28 protein biomarkers is shown in ESM Fig. 4. The results of PPI network and enrichment analyses are shown in the ESM Fig. 5, 6.

We performed several sensitivity analyses regarding incident type 2 diabetes as the outcome. After adjusting for diabetes-related lifestyle factors, associations for LPL and FGF21 lost significance (p_FDR=0.054) (ESM Table 3). All associations remained statistically significant after consideration of 206 deaths as a competing risk and after adjusting for use of lipid-lowering medication at baseline. When we excluded the 262 participants who only had questionnaire-based information regarding the development of incident type 2 diabetes, PON3 was not significantly associated with incident type 2 diabetes (p_FDR=0.658) (ESM Table 4). Among normoglycaemic participants at baseline, the effect estimates of the four incident type 2 diabetes-related proteins went in the same direction but lost statistical significance, most likely due to power limitations (ESM Table 5).

Associations of proteins with continuous outcomes

The identified 28 diabetes-related proteins were further included in the analysis of traits of blood glucose and insulin resistance and secretion. The trajectories of all six traits are shown in Fig. 3 stratified by diabetes status by the end of follow-up. We found six to 21 proteins associated with fasting glucose, 2 h glucose, fasting insulin, HOMA-IR, HOMA-B or HbA_{1c} (Fig. 4 and ESM Tables 6, 7). IGFBP2, LPL, hepatocyte growth factor and IGF binding protein 1 were found to be associated with all traits either cross-sectionally or longitudinally. In the longitudinal results, of the four incident type 2 diabetes-related proteins, LPL was associated with fasting glucose and HbA_{1c}, while IGFBP2 and PON3 were associated with both glucose- and insulin-related traits.

In a sensitivity analysis adjusting for glucose-lowering medication intake, similar associations between proteins and continuous outcomes were observed (ESM Tables 8, 9), except for the cross-sectional results of fasting glucose. Here only one of 12 proteins (hydroxyacid oxidase 1) remained statistically significant. IL-1 receptor type 2 (IL-1RT2) lost statistical significance with HOMA-IR cross-sectionally after adjusting for glucose-lowering medication.

Comparison of identified diabetes-related markers with previous studies

We assessed the overlap between our identified proteins and previously reported diabetes-related markers by searching the Human Diabetes Proteome Project published in 2014 [25] and additional epidemiological publications after 2013 [3, 4, 12, 13, 26, 27]. The three searching strategies employed in the present study and the summarised results can be found in ESM Table 10.

In summary, we observed that five of our identified proteins, namely IL-17D, IL-18 receptor 1 (IL-18R1), carbonic anhydrase-5A (CA5A), IL-1RT2 and matrix extracellular phosphoglycoprotein (MEPE), have not been previously reported to be associated with either prevalent or incident type 2 diabetes or prediabetes.

Causal effects of top proteins on type 2 diabetes and continuous traits

We found 177 *cis*-acting genetic IVs for our top 28 diabetes-related proteins from previously published GWAS, and examined possible causal effects (ESM Table 11). LPL was the only protein for which we observed a statistically significant causal effect on type 2 diabetes (Wald ratio, $b = -0.3564$; p value = 7.23×10^{-7}) and fasting insulin (Wald ratio, $b = -0.0752$; p value = 0.0027). Regarding MR analysis on other traits, we found no evidence of a causal association between the IVs and respective outcomes after adjusting for multiple testing.

Prediction of incident type 2 diabetes

The top 14 priority-Lasso-selected proteins (selection frequency >20%) were added consecutively to the basic FORS model. The best set of predictors for incident type 2 diabetes consisted of the top 12 proteins (IGFBP2, IL-18, IL-17D, complement component C1q receptor [CD93], V-set and immunoglobulin domain-containing protein 2 [VSIG2], IL-1RT2, LPL, CUB domain-containing protein 1 [CDCP1], vascular endothelial growth factor D [VEGFD], PON3, C-C motif chemokine 4 [CCL4] and tartrate-resistant acid phosphatase type 5 [TR-AP]) (ESM Table 12). The mean AUC value of this set of predictors for incident type 2 diabetes was 0.7699, which was 2.9% (ΔAUC [95% CI]=0.0219 [0.0052, 0.0624]) higher than the corresponding AUC value of the FORS model (0.7480). Moreover, IGFBP2 was the most important protein and was selected in 847 over 1000 repeats (ESM Table 13).

Discussion

This study provides a comprehensive large-scale analysis of proteomics data, identifying novel biomarkers and replicating previously identified proteins possibly involved in the

Table 1 Baseline characteristics of the study population (n=1300)

Characteristic	Normoglycaemia (n=840)	Prediabetes (n=344)	Newly diagnosed type 2 diabetes (n=116)
Age (years)	63.4±5.5	65.0±5.2	65.0±5.4
Male	401 (47.7)	203 (59.0)	71 (61.2)
Parental diabetes			
Yes	178 (21.2)	85 (24.7)	48 (41.4)
No	494 (58.8)	186 (54.1)	42 (36.2)
Unknown	168 (20.0)	73 (21.2)	26 (22.4)
BMI (kg/m^2)	27.6±4.1	29.6±4.1	30.0±3.9
Systolic BP (mmHg)	131.6±18.9	140.2±19.1	145.9±22.1
Diastolic BP (mmHg)	78.8±10.0	81.8±10.4	81.6±10.1
Waist circumference (cm)	93.2±11.2	99.3±10.4	101.8±10.8
HDL-cholesterol (mmol/l)	1.56±0.42	1.43±0.40	1.35±0.42
Triglycerides (mmol/l)	1.21 (0.90, 1.65)	1.42 (1.05, 2.00)	1.61 (1.22, 2.17)
Fasting glucose (mmol/l)	5.28 (5.06, 5.61)	6.11 (5.61, 6.33)	7.11 (6.33, 7.83)
2 h glucose (mmol/l)[a]	5.72 (4.83, 6.56)	8.39 (7.10, 9.50)	12.39 (11.17, 14.00)
Fasting insulin (pmol/l)[b]	52.2 (37.8, 73.8)	74.7 (55.8, 109.8)	86.4 (54.5, 125.1)
HOMA-IR[b]	2.06 (1.49, 2.95)	3.36 (2.36, 5.00)	4.69 (2.62, 7.58)
HOMA-B[b]	100.0 (72.0, 138.4)	105.7 (77.7, 140.0)	77.0 (51.3, 117.7)
HbA_{1c} (mmol/mol)[c]	38.0 (34.0, 40.0)	38.0 (36.0, 41.0)	42.0 (39.0, 46.0)
HbA_{1c} (%)[c]	5.6 (5.3, 5.8)	5.6 (5.4, 5.9)	6.0 (5.7, 6.4)
Medication use			
Antihypertensive medication use	237 (28.2)	164 (47.7)	47 (40.5)
Statin use	73 (8.7)	35 (10.2)	13 (11.2)
Lipid-lowering drug treatment	87 (10.4)	39 (11.3)	18 (15.5)
Physically active[d]	396 (47.2)	129 (37.6)	38 (32.8)
Smoking			
Never smoker	410 (48.8)	162 (47.1)	45 (38.8)
Former smoker	309 (36.8)	143 (41.6)	51 (44.0)
Current smoker	121 (14.4)	39 (11.3)	19 (16.4)
Alcohol intake (g/day)[d]	6.29 (0, 22.86)	13.20 (0, 26.80)	6.60 (0, 25.71)
Meat consumption (frequency/day)[d]	0.50 (0.14, 0.50)	0.50 (0.14, 0.50)	0.50 (0.14, 0.50)
Whole-grain bread/muesli consumption (frequency/day)[d]	1.00 (0.50, 1.07)	1.00 (0.21, 1.03)	0.61 (0.14, 1.00)
Coffee consumption (cups/day)[d]	2 (2, 4)	2 (1, 4)	2 (1, 4)

Continuous variables are presented as mean±SD for normally distributed data and as median (25th, 75th) for data not normally distributed. Categorical variables are presented as n (%)

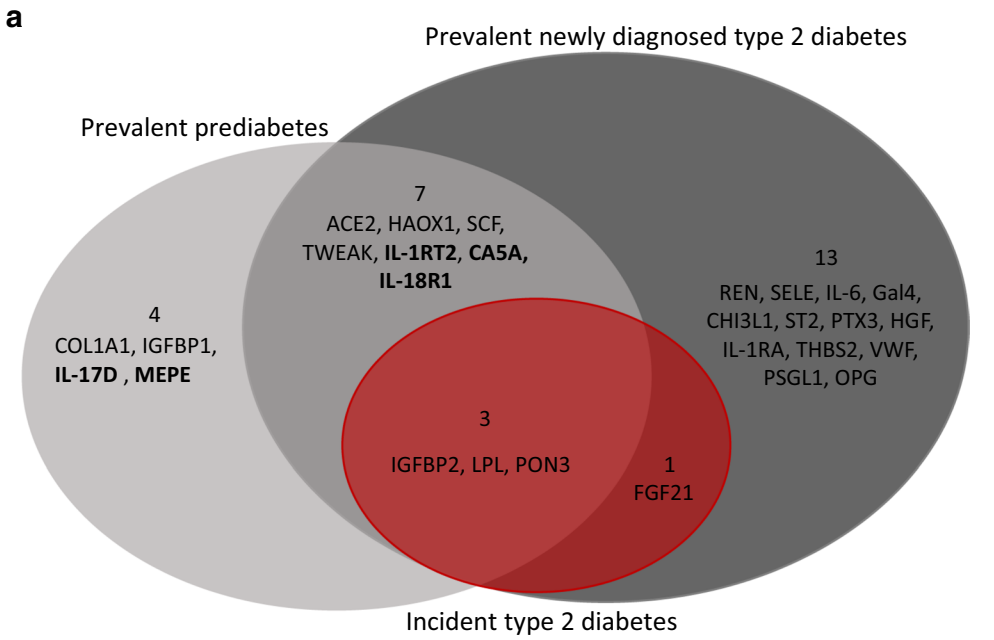
[a]Data were calculated in 1297 participants at baseline (840 normoglycaemia, 344 prediabetes and 113 type 2 diabetes)

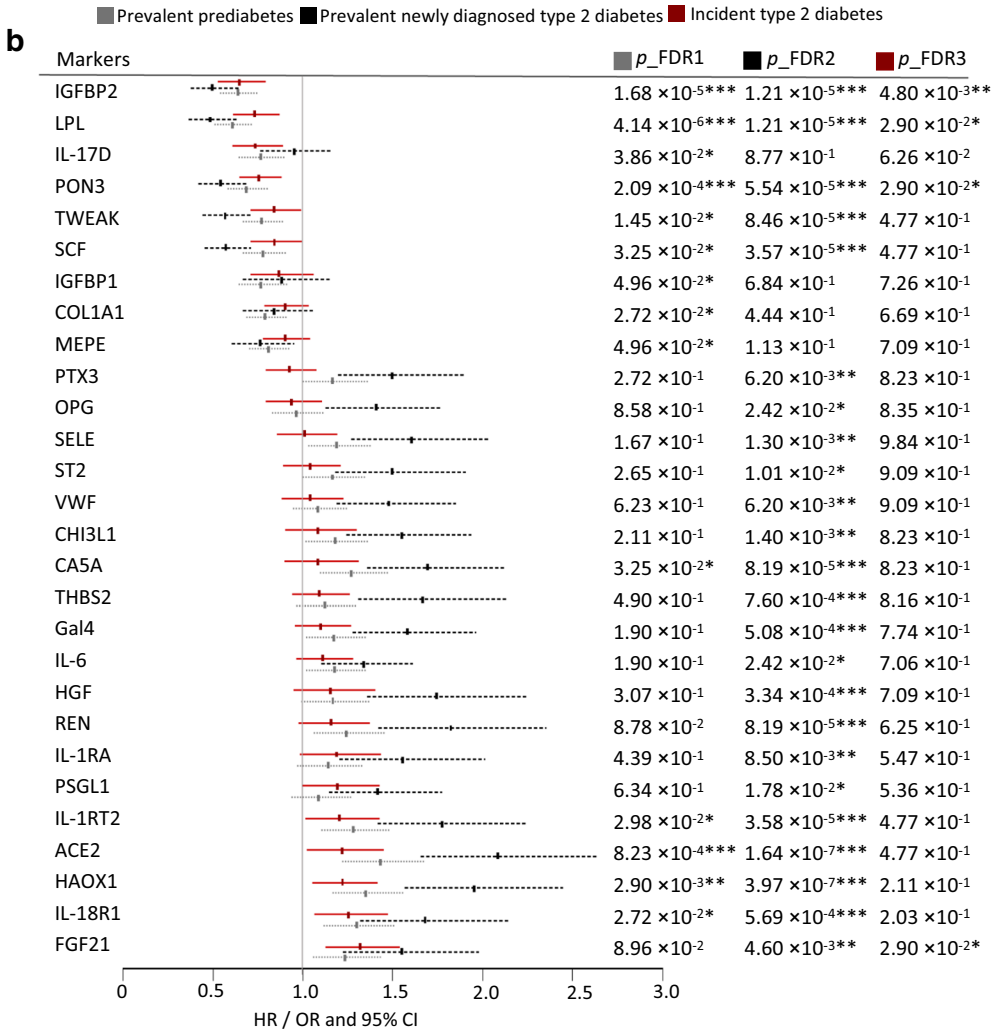
[b]Data were calculated in 1296 participants at baseline (839 normoglycaemia, 342 prediabetes and 115 type 2 diabetes)

[c]Data were calculated in 1299 participants at baseline (840 normoglycaemia, 343 prediabetes and 116 type 2 diabetes)

[d]Data were calculated in 1298 participants at baseline (839 normoglycaemia, 343 prediabetes and 116 type 2 diabetes)

pathophysiology of derangements in glucose metabolism and type 2 diabetes development. Specifically, we identified 28 diabetes-related proteins and five of them, including IL-17D, IL-18R1, CA5A, IL-1RT2 and MEPE, were reported for the first time. Four proteins were found to be associated with incident type 2 diabetes (IGFBP2, LPL, PON3 and FGF21). Longitudinally, these biomarkers were associated with changes in glucose-related traits only (LPL), or both glucose-related traits and insulin-related traits (IGFBP2 and PON3). MR analysis provided suggestive evidence for a causal relationship between LPL and type 2 diabetes. The combination


a
Prevalent newly diagnosed type 2 diabetes
Prevalent prediabetes
7
ACE2, HAOX1, SCF,
TWEAK, IL-1RT2, CA5A,
IL-18R1
13
REN, SELE, IL-6, Gal4,
CHI3L1, ST2, PTX3, HGF,
IL-1RA, THBS2, VWF,
PSGL1, OPG
4
COL1A1, IGFBP1,
IL-17D , MEPE
3
IGFBP2, LPL, PON3
1
FGF21
Incident type 2 diabetes


Prevalent prediabetes
Prevalent newly diagnosed type 2 diabetes
Incident type 2 diabetes
b
Markers
p_FDR1
p_FDR2
p_FDR3
IGFBP2 1.68 ×10-5*** 1.21 ×10-5*** 4.80 ×10-3**
LPL 4.14 ×10-6*** 1.21 ×10-5*** 2.90 ×10-2*
IL-17D 3.86 ×10-2* 8.77 ×10-1 6.26 ×10-2
PON3 2.09 ×10-4*** 5.54 ×10-5*** 2.90 ×10-2*
TWEAK 1.45 ×10-2* 8.46 ×10-5*** 4.77 ×10-1
SCF 3.25 ×10-2* 3.57 ×10-5*** 4.77 ×10-1
IGFBP1 4.96 ×10-2* 6.84 ×10-1 7.26 ×10-1
COL1A1 2.72 ×10-2* 4.44 ×10-1 6.69 ×10-1
MEPE 4.96 ×10-2* 1.13 ×10-1 7.09 ×10-1
PTX3 2.72 ×10-1 6.20 ×10-3** 8.23 ×10-1
OPG 8.58 ×10-1 2.42 ×10-2* 8.35 ×10-1
SELE 1.67 ×10-1 1.30 ×10-3** 9.84 ×10-1
ST2 2.65 ×10-1 1.01 ×10-2* 9.09 ×10-1
VWF 6.23 ×10-1 6.20 ×10-3** 9.09 ×10-1
CHI3L1 2.11 ×10-1 1.40 ×10-3** 8.23 ×10-1
CA5A 3.25 ×10-2* 8.19 ×10-5*** 8.23 ×10-1
THBS2 4.90 ×10-1 7.60 ×10-4*** 8.16 ×10-1
Gal4 1.90 ×10-1 5.08 ×10-4*** 7.74 ×10-1
IL-6 1.90 ×10-1 2.42 ×10-2* 7.06 ×10-1
HGF 3.07 ×10-1 3.34 ×10-4*** 7.09 ×10-1
REN 8.78 ×10-2 8.19 ×10-5*** 6.25 ×10-1
IL-1RA 4.39 ×10-1 8.50 ×10-3** 5.47 ×10-1
PSGL1 6.34 ×10-1 1.78 ×10-2* 5.36 ×10-1
IL-1RT2 2.98 ×10-2* 3.58 ×10-5*** 4.77 ×10-1
ACE2 8.23 ×10-4*** 1.64 ×10-7*** 4.77 ×10-1
HAOX1 2.90 ×10-3** 3.97 ×10-7*** 2.11 ×10-1
IL-18R1 2.72 ×10-2* 5.69 ×10-4*** 2.03 ×10-1
FGF21 8.96 ×10-2 4.60 ×10-3** 2.90 ×10-2*
0
0.5
1.0
1.5
2.0
2.5
3.0
HR / OR and 95% CI

◂**Fig. 2** Twenty-eight identified proteins and their associations with prevalent prediabetes, prevalent newly diagnosed diabetes and incident type 2 diabetes. (**a**) The overlap between proteins associated with prevalent prediabetes (light grey), prevalent newly diagnosed type 2 diabetes (dark grey) and incident type 2 diabetes (red), respectively. The novel diabetes-related proteins are marked in bold. (**b**) Forest plot summarising the results of the main analyses in model 2. Effect estimates have been calculated per 1 SD increase in NPX values on a $\log_2$ scale. The grey dashed line, the black dashed line and the red line represent OR and 95% CI of prevalent prediabetes and prevalent newly diagnosed type 2 diabetes, and the HR and 95% CI of incident type 2 diabetes, respectively. *p*_FDR1, *p*_FDR2 and *p*_FDR3 present the *p* values for prevalent prediabetes, prevalent newly diagnosed type 2 diabetes and incident type 2 diabetes after controlling for the FDR, respectively. Model 2 was adjusted for age, sex, parental history of diabetes, systolic blood pressure, BMI, HDL-cholesterol, triglycerides and waist circumference. Protein biomarkers are sorted by strength of association for incident type 2 diabetes. *$p<0.05$, **$p<0.01$, ***$p<0.001$. CHI3L1, chitinase-3-like protein 1; COL1A1, collagen alpha-1(I) chain; HAOX1, hydroxyacid oxidase 1; HGF, hepatocyte growth factor; IGFBP1, IGF binding protein 1; OPG, osteoprotegerin; PSGL1, P-selectin glycoprotein ligand 1; PTX3, pentraxin-related protein 3; REN, renin; SCF, stem cell factor; SELE, E-selectin; ST2, ST2 protein; THBS2, thrombospondin-2; TWEAK, tumor necrosis factor (Ligand) superfamily member 12; VWF, von Willebrand factor

of 12 selected proteins yielded the best improvement in type 2 diabetes prediction beyond classical diabetes risk factors.

Novel markers for prevalent prediabetes and type 2 diabetes

Among the five new candidates related to prediabetes and/or type 2 diabetes, three (IL-1RT2, CA5A and IL-18R1) were positively associated with prevalent prediabetes and type 2 diabetes. IL-1RT2 is an IL-1 family receptor, involved in the regulation of immune and inflammatory responses. It was included in type 2 diabetes-associated coexpression genes (module) enriched for IL-1-related genes [28]; however, our study is the only epidemiological study connecting IL-1RT2 to type 2 diabetes and implicating IL-1RT2 to be involved in altered glucose metabolism and insulin resistance. CA5A is a ubiquitous zinc metalloenzyme, playing a vital role in various biosynthetic processes such as gluconeogenesis and lipogenesis [29], and increases in carbonic anhydrase (CA) activity have been reported to increase the production of hepatic glucose in type 2 diabetes [30]. In line with this, our study showed that higher CA5A was associated with a higher level of fasting glucose. IL-18R1 is a subunit of the proinflammatory factor IL-18 receptor and exists in membrane-bound and soluble forms. Elevated levels of IL-18 have been linked to an increased risk of type 2 diabetes [19, 31] and previous epidemiological studies have reported positive associations of plasma IL-18R1 with the metabolic syndrome [32] and obesity [33]. Similarly, Mahdi et al observed an association between the expression of *IL18R1* and HbA_{1c} levels among people without diabetes [28].

In the present study, IL-17D and MEPE were inversely associated with prevalent prediabetes. IL-17D, a member of the IL-17 cytokine family highly expressed in the brain and skeletal muscle, has previously been associated with autoimmune and inflammatory diseases [34]. Moreover, DNA methylation at *IL17D* was associated with maternal early-pregnancy glucose concentrations [35]. Of note, IL-17D is required for maintaining intestinal homeostasis so reduced IL-17D levels could be related to dysbiosis and higher prediabetes risk [36]. MEPE is involved in the formation of the extracellular matrix of bone and the renal regulation of bone mineralisation including phosphate homeostasis [37]. New evidence has implicated this protein in pathways related to diabetes and obesity [38].

Known markers for incident type 2 diabetes

In agreement with previous reports, we found associations of LPL, IGFBP2, FGF21 and PON3 with incident type 2 diabetes. LPL is an important enzyme in triglyceride metabolism and has been shown to impact type 2 diabetes development via improvements of insulin resistance and regulation of dyslipidaemia [39, 40]. Our MR analysis revealed a suggestive causal protective association between LPL and type 2 diabetes, with a consistent directionality also for an inverse effect on fasting insulin. Previous studies support both our observational and MR analysis findings [3]. The inverse association between IGFBP2 and incident type 2 diabetes observed in our study aligns with evidence from epidemiological studies [14, 41] and experimental work in mice [42]. FGF21 has been reported to be an important endocrine factor, regulating glucose and lipid metabolism, increasing insulin sensitivity and improving islet beta cell secretion and proliferation, with the potential to be a target for diabetes treatment [43–45]. However, in line with the present study, elevated FGF21 concentrations were observed in patients with diabetes or obesity, possibly to compensate insulin deficiency [43, 46, 47]. PON3 is bound to HDL-cholesterol in the circulation and is closely related to insulin resistance, lipid metabolism and obesity [48]. Previous epidemiological reports confirm our results of an inverse association with prevalent/incident type 2 diabetes [3, 12].

Prediction of incident diabetes

The addition of the top 12 selected proteins yielded the highest improved predictive performance with a ΔAUC of 0.0219, but, of note, the addition of the top nine proteins had a ΔAUC of 0.0218 which was only marginally lower. When we excluded fasting glucose from the

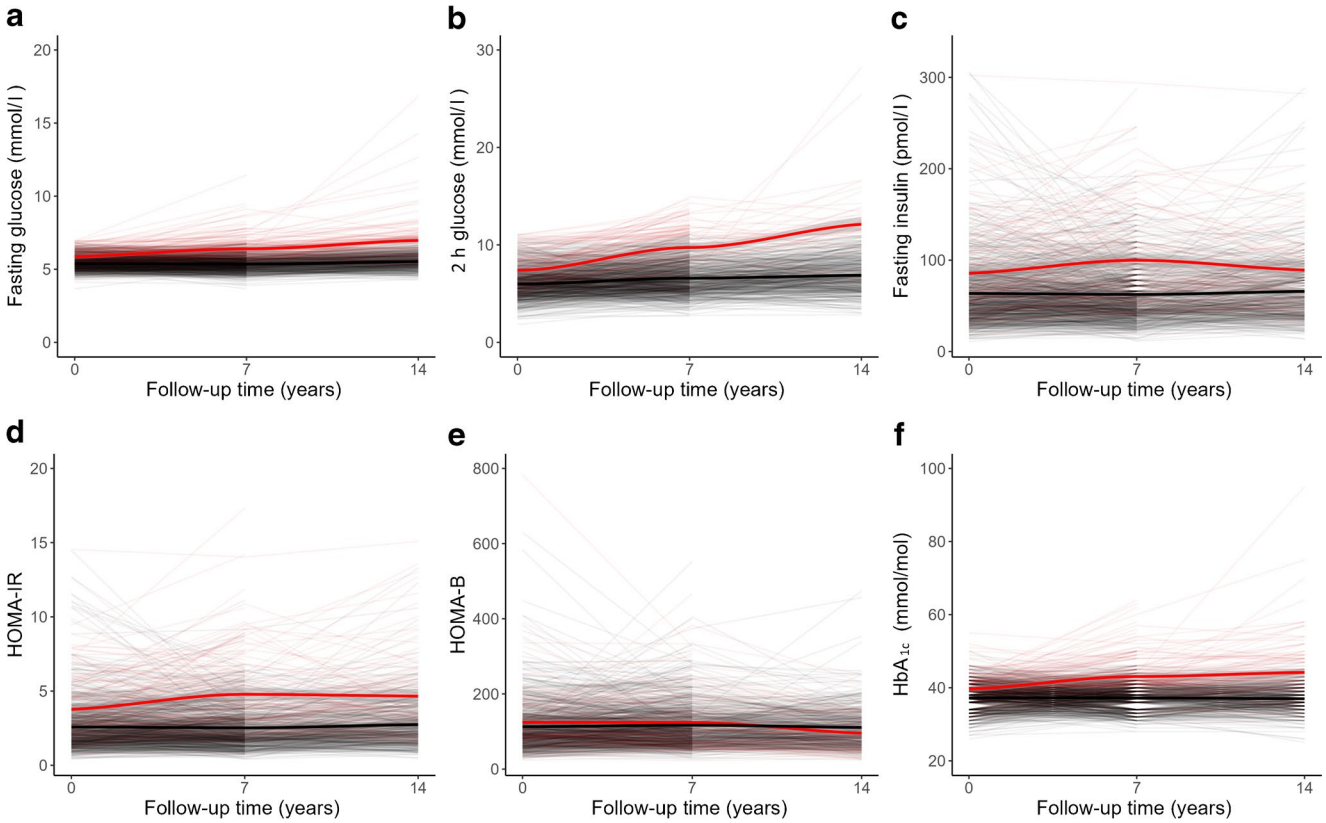


Fig. 3 Descriptive figure showing the trajectories of traits of blood glucose and insulin resistance and secretion during 14 years of follow-up grouped by incident type 2 diabetes. The red line and the black line represent participants with and without type 2 diabetes, respectively

reference model, we found a ΔAUC of 0.0477 (AUC of the corresponding basic model was 0.7040). Previous studies have tried to improve prediction models for type 2 diabetes using biomarkers [5, 9, 11, 12, 14, 15]; however, only three of them showed improvement in discrimination by adding selected proteins to a clinical model, and the magnitude of model improvement was moderate with ΔAUC values ranging from 0.012 to 0.034 [11, 14, 15]. The direct comparisons of ΔAUC are restricted due to differences in baseline clinical models, available biomarker panels and analytical approaches. Of note, the ΔAUC was the lowest (0.012) in the study of Salomaa et al [15] which also included fasting glucose in the basic model. Moreover, when Huth et al [11] added HbA_{1c} to the basic non-invasive risk factor model, their extended protein model showed a ΔAUC of 0.005. Up to now, there is still little consensus regarding specific sets of biomarkers that would improve diabetes prediction. Therefore, further prospective studies with larger samples are needed to validate the protein set identified in the present analysis.

Study strengths and limitations

The strengths of the present study include the examination of a large number of proteins by PEA technology with regard to type 2 diabetes and related traits in a population-based study population. The availability of OGTT data at baseline and at up to two follow-up examinations characterises well changes in a wide range of diabetes-related outcomes/traits. Finally, we were able to evaluate the causal relationship using publicly available data on genetic associations of identified proteins with type 2 diabetes and related traits.

The present study also has some limitations. First, the PEA approach provided only relative and not absolute protein concentrations, which, however, does not affect the reported associations. Second, proteomics measurements were only performed at baseline, precluding us from taking into consideration the impact of changes in protein concentrations on the progression towards type 2 diabetes. Third, we lack comprehensive dietary data to capture overall diet quality. Finally, although this study used an internal

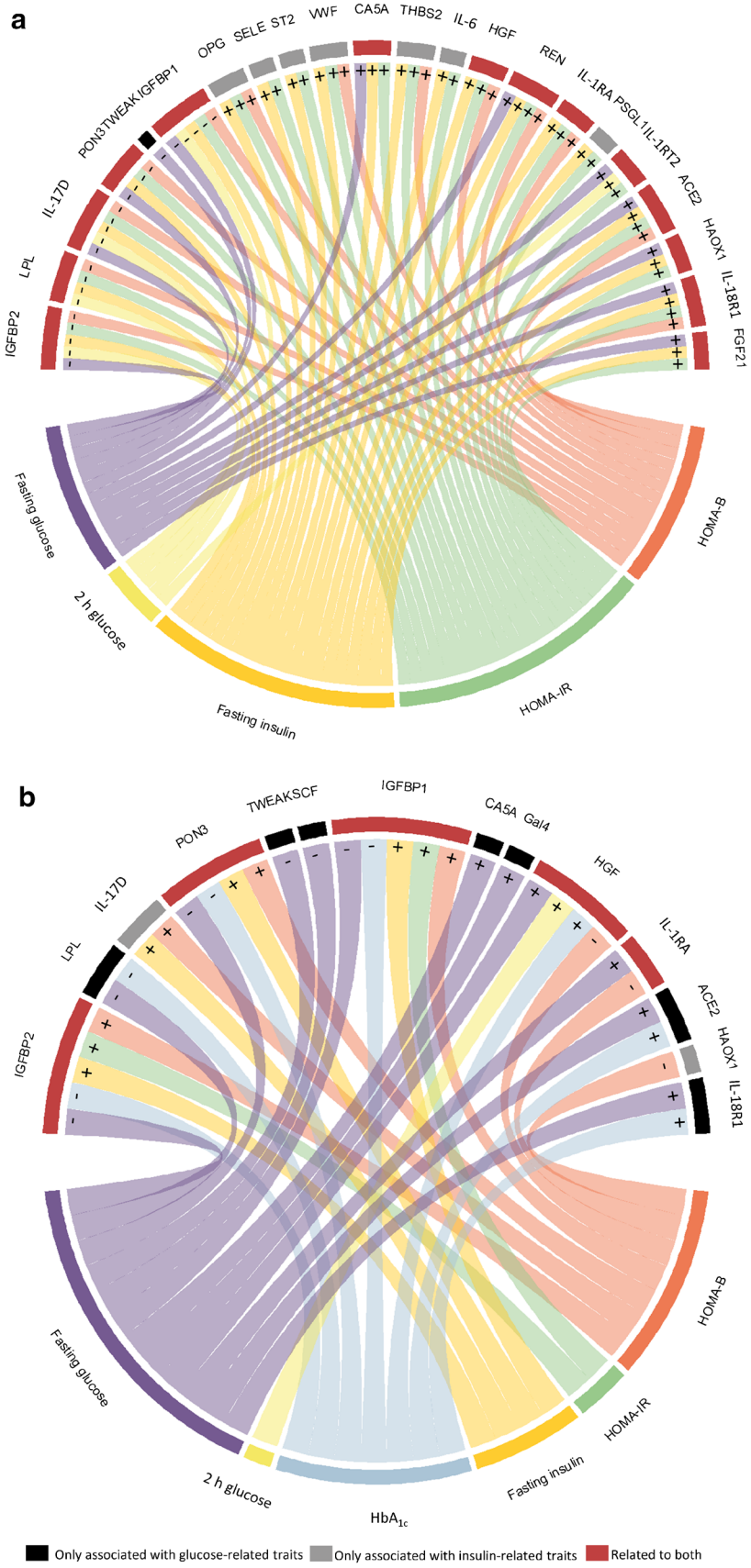


Fig. 4 Chord diagram showing significant associations of traits of blood glucose and insulin resistance and secretion with circulating protein levels. (**a**) Significant proteins cross-sectionally associated with traits of blood glucose and insulin resistance and secretion. (**b**) Significant proteins longitudinally associated with traits of blood glucose and insulin resistance and secretion. The '+' indicates positive association, while '−' indicates inverse association. HAOX1, hydroxyacid oxidase 1; HGF, hepatocyte growth factor; OPG, osteoprotegerin; PSGL1, P-selectin glycoprotein ligand 1; REN, renin; SCF, stem cell factor; SELE, E-selectin; ST2, ST2 protein; THBS2, thrombospondin-2; TWEAK, tumor necrosis factor (Ligand) superfamily member 12; VWF, von Willebrand factor

cross-validation method to evaluate the predictive performance, external replication studies are needed to validate our findings regarding the prediction of type 2 diabetes and to confirm the relevance of novel protein candidates for type 2 diabetes.

In summary, we identified five novel candidates possibly involved in the pathophysiology of type 2 diabetes and replicated previously reported associations with type 2 diabetes. Our results provide new insight into the aetiological roles of plasma proteins in glucose and insulin metabolism and type 2 diabetes. Further characterisation of the novel biomarkers identified in this study offers the potential to help us uncover new mechanisms that lead to type 2 diabetes and discover new drug targets.

Supplementary Information The online version contains peer-reviewed but unedited supplementary material available at https://doi.org/10.1007/s00125-023-05943-2.

Acknowledgements We thank the UK Biobank Pharma Proteomics Project (UKB-PPP, AMS Application 65851) for providing access to a selection of proteo-genomic association statistics—with special thanks to C. D. Whelan (Translation Sciences, Research & Development, Biogen Inc., Cambridge, MA, USA) for leading the study. Some of the data were presented as an abstract at the IDEG (International Diabetes Epidemiology Group) conference and the IDF (International Diabetes Federation) meeting in 2022.

Data availability Data from this KORA study are not publicly available because the data are subject to national data protection laws, and restrictions were imposed by the ethics committee of the Bavarian Chamber of Physicians to ensure data privacy of the study participants. However, data are available on request to researchers through a project agreement from KORA (http://epi.helmholtz-muenchen.de/kora-gen/). Requests should be sent to kora.passt@helmholtz-muenchen.de and are subject to approval by the KORA board.

Funding HL was supported by a scholarship under the State Scholarship Fund by the China Scholarship Council (File No. 201906380066). The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Data collection in the KORA study is done in cooperation with the University Hospital of Augsburg. The present study was supported by Helmholtz Institute for Metabolic, Obesity and Vascular Research – Project Initiative 2018 (HI-MAG). The German Diabetes Center is funded by the German Federal Ministry of Health (Berlin, Germany) and the Ministry of Culture and Science of the state North Rhine-Westphalia (Düsseldorf, Germany) and receives additional funding from the German Federal Ministry of Education and Research (BMBF) through the German Center for Diabetes Research (DZD e.V.). The UK Biobank Pharma Proteomics Project was funded by the research and development leadership teams at the 13 participating UKB-PPP member companies (Alnylam Pharmaceuticals, Amgen, AstraZeneca, Biogen, Bristol-Myers Squibb, Calico, Genentech, Glaxo Smith Klein, Janssen Pharmaceuticals, Novo Nordisk, Pfizer, Regeneron and Takeda).

Authors' relationships and activities WR reports receiving consulting fees for attending educational sessions or advisory boards from AstraZeneca, Boehringer Ingelheim and Novo Nordisk and institutional research grants from Novo Nordisk outside of the topic of the current paper. CH is member of the editorial board of *Diabetologia*. BBS is an employee of Biogen. The authors declare that there are no other relationships or activities that might bias, or be perceived to bias, their work.

Contribution statement HL and BT conceptualised the research question. HL drafted the analysis plan, performed the statistical analysis and wrote the manuscript with guidance from JN and BT. AB and AH provided statistical analysis advice. BBS did genetic association analyses with Olink Proteomics. A. Petrera, WR, CH, SMH, A. Peters and BT contributed data. All authors were involved in the review and final approval of the manuscript. HL is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. Mathers CD, Loncar D (2006) Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 3(11):e442. https://doi.org/10.1371/journal.pmed.0030442
2. Gudmundsdottir V, Zaghlool SB, Emilsson V et al (2020) Circulating protein signatures and causal candidates for type 2 diabetes. Diabetes 69(8):1843–1853. https://doi.org/10.2337/db19-1070
3. Beijer K, Nowak C, Sundstrom J, Arnlov J, Fall T, Lind L (2019) In search of causal pathways in diabetes: a study using proteomics and genotyping data from a cross-sectional study. Diabetologia 62(11):1998–2006. https://doi.org/10.1007/s00125-019-4960-8
4. Ferreira JP, Pizard A, Machu JL et al (2020) Plasma protein biomarkers and their association with mutually exclusive cardiovascular phenotypes: the FIBRO-TARGETS case-control analyses. Clin Res Cardiol 109(1):22–33. https://doi.org/10.1007/s00392-019-01480-4
5. Elhadad MA, Jonasson C, Huth C et al (2020) Deciphering the plasma proteome of type 2 diabetes. Diabetes 69(12):2766–2778. https://doi.org/10.2337/db20-0296
6. Tromp J, Voors AA, Sharma A et al (2020) Distinct pathological pathways in patients with heart failure and diabetes. JACC Heart Fail 8(3):234–242. https://doi.org/10.1016/j.jchf.2019.11.005
7. von Toerne C, Huth C, de Las Heras Gala T et al (2016) MASP1, THBS1, GPLD1 and ApoA-IV are novel biomarkers associated with prediabetes: the KORA F4 study. Diabetologia 59(9):1882–1892. https://doi.org/10.1007/s00125-016-4024-2
8. Belongie KJ, Ferrannini E, Johnson K, Andrade-Gordon P, Hansen MK, Petrie JR (2017) Identification of novel biomarkers to monitor beta-cell function and enable early detection of type 2 diabetes risk. PLoS One 12(8):e0182932. https://doi.org/10.1371/journal.pone.0182932
9. Nowak C, Sundstrom J, Gustafsson S et al (2016) Protein biomarkers for insulin resistance and type 2 diabetes risk in two large community cohorts. Diabetes 65(1):276–284. https://doi.org/10.2337/db15-0881
10. Curran AM, Scott-Boyer MP, Kaput J et al (2018) A proteomic signature that reflects pancreatic beta-cell function. PLoS One 13(8):e0202727. https://doi.org/10.1371/journal.pone.0202727
11. Huth C, von Toerne C, Schederecker F et al (2019) Protein markers and risk of type 2 diabetes and prediabetes: a targeted proteomics approach in the KORA F4/FF4 study. Eur J Epidemiol 34(4):409–422. https://doi.org/10.1007/s10654-018-0475-8
12. Molvin J, Pareek M, Jujic A et al (2019) Using a targeted proteomics chip to explore pathophysiological pathways for incident diabetes- the Malmo Preventive Project. Sci Rep 9(1):272. https://doi.org/10.1038/s41598-018-36512-y
13. Ramne S, Drake I, Ericson U et al (2020) Identification of inflammatory and disease-associated plasma proteins that

associate with intake of added sugar and sugar-sweetened beverages and their role in type 2 diabetes risk. Nutrients 12(10):3192. https://doi.org/10.3390/nu12103129
14. Thorand B, Zierer A, Buyukozkan M et al (2021) A panel of 6 biomarkers significantly improves the prediction of type 2 diabetes in the MONICA/KORA study population. J Clin Endocrinol Metab 106(4):e1647–e1659. https://doi.org/10.1210/clinem/dgaa953
15. Salomaa V, Havulinna A, Saarela O et al (2010) Thirty-one novel biomarkers as predictors for clinically incident diabetes. PLoS One 5(4):e10100. https://doi.org/10.1371/journal.pone.0010100
16. Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB Sr (2007) Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. Arch Intern Med 167(10):1068–1074. https://doi.org/10.1001/archinte.167.10.1068
17 Rathmann W, Haastert B, Icks A et al (2003) High prevalence of undiagnosed diabetes mellitus in Southern Germany: target populations for efficient screening. The KORA survey 2000. Diabetologia 46(2):182–189. https://doi.org/10.1007/s00125-002-1025-0
18. Petrera A, von Toerne C, Behler J et al (2021) Multiplatform approach for plasma proteomics: complementarity of olink proximity extension assay technology to mass spectrometry-based protein profiling. J Proteome Res 20(1):751–762. https://doi.org/10.1021/acs.jproteome.0c00641
19. Thorand B, Kolb H, Baumert J et al (2005) Elevated levels of interleukin-18 predict the development of type 2 diabetes: results from the MONICA/KORA Augsburg Study, 1984–2002. Diabetes 54(10):2932–2938. https://doi.org/10.2337/diabetes.54.10.2932
20. Grimm KJ, Ram N, Estabrook R (2016) Growth modeling: structural equation and multilevel modeling approaches, 1st edn. Guilford Publications, New York
21. Teumer A (2018) Common methods for performing mendelian randomization. Front Cardiovasc Med 5:51. https://doi.org/10.3389/fcvm.2018.00051
22. Bowden J, Davey Smith G, Burgess S (2015) Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol 44(2):512–525. https://doi.org/10.1093/ije/dyv080
23. Klau S, Jurinovic V, Hornung R, Herold T, Boulesteix AL (2018) Priority-Lasso: a simple hierarchical approach to the prediction of clinical outcome using multi-omics data. BMC Bioinformatics 19(1):322. https://doi.org/10.1186/s12859-018-2344-6
24. Pencina MJ, D'Agostino RB (2004) Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. Stat Med 23(13):2109–2123. https://doi.org/10.1002/sim.1802
25. Schvartz D, Bergsten P, Baek KH, Barba De La Rosa A, Cantley J, Dayon L (2015) The Human Diabetes Proteome Project (HDPP): the 2014 update. Transl Proteom 8–9:1–7. https://doi.org/10.1016/j.trprot.2015.03.001
26. Diaz-Lopez A, Chacon MR, Bullo M et al (2013) Serum sTWEAK concentrations and risk of developing type 2 diabetes in a high cardiovascular risk population: a nested case-control study. J Clin Endocrinol Metab 98(8):3482–3490. https://doi.org/10.1210/jc.2013-1848
27. Lin YH, Zhang RC, Hou LB et al (2016) Distribution and clinical association of plasma soluble ST2 during the development of type 2 diabetes. Diabetes Res Clin Pract 118:140–145. https://doi.org/10.1016/j.diabres.2016.06.006
28. Mahdi T, Hänzelmann S, Salehi A et al (2012) Secreted frizzled-related protein 4 reduces insulin secretion and is overexpressed in type 2 diabetes. Cell Metab 16(5):625–633. https://doi.org/10.1016/j.cmet.2012.10.009
29. Queen A, Khan P, Azam A, Hassan MI (2018) Understanding the role and mechanism of carbonic anhydrase V in obesity and its therapeutic implications. Curr Protein Pept Sci 19(9):909–923. https://doi.org/10.2174/1389203718666170810145436
30. Ismail IS (2018) The role of carbonic anhydrase in hepatic glucose production. Curr Diabetes Rev 14(2):108–112. https://doi.org/10.2174/1573399812666161214122351
31. Bankul A, Mitra P, Suri S et al (2023) Increased serum IL-18 levels and IL-18R expression in newly diagnosed type 2 diabetes mellitus. Minerva Endocrinol 48(1):35–41. https://doi.org/10.23736/S0391-1977.20.03271-X
32. Van Alsten SC, Rabkin CS, Sawada N et al (2020) Metabolic syndrome, physical activity, and inflammation: a cross-sectional analysis of 110 circulating biomarkers in Japanese adults. Cancer Epidemiol Biomarkers Prev 29(8):1639–1646. https://doi.org/10.1158/1055-9965.EPI-19-1513
33. Ponce-de-Leon M, Linseisen J, Peters A et al (2022) Novel associations between inflammation-related proteins and adiposity: a targeted proteomics approach across four population-based studies. Transl Res 242:93–104. https://doi.org/10.1016/j.trsl.2021.11.004
34. Matsuzaki G, Umemura M (2018) Interleukin-17 family cytokines in protective immunity against infections: role of hematopoietic cell-derived and non-hematopoietic cell-derived interleukin-17s. Microbiol Immunol 62(1):1–13. https://doi.org/10.1111/1348-0421.12560
35. Geurtsen ML, Jaddoe VWV, Gaillard R, Felix JF (2020) Associations of maternal early-pregnancy blood glucose and insulin concentrations with DNA methylation in newborns. Clin Epigenetics 12(1):134. https://doi.org/10.1186/s13148-020-00924-3
36. Huang J, Lee HY, Zhao X et al (2021) Interleukin-17D regulates group 3 innate lymphoid cell function through its receptor CD93. Immunity 54(4):673–686. https://doi.org/10.1016/j.immuni.2021.03.018. (e674)
37. Confavreux CB, Levine RL, Karsenty G (2009) A paradigm of integrative physiology, the crosstalk between bone and energy metabolisms. Mol Cell Endocrinol 310(1–2):21–29. https://doi.org/10.1016/j.mce.2009.04.004
38. Rowe PS (2012) Regulation of bone-renal mineral and energy metabolism: the PHEX, FGF23, DMP1, MEPE ASARM pathway. Crit Rev Eukaryot Gene Expr 22(1):61–86. https://doi.org/10.1615/critreveukargeneexpr.v22.i1.50
39. Lotta LA, Stewart ID, Sharp SJ et al (2018) Association of genetically enhanced lipoprotein lipase-mediated lipolysis and low-density lipoprotein cholesterol-lowering alleles with risk of coronary disease and type 2 diabetes. JAMA Cardiol 3(10):957–966. https://doi.org/10.1001/jamacardio.2018.2866
40. Cho YS, Go MJ, Han HR et al (2008) Association of lipoprotein lipase (LPL) single nucleotide polymorphisms with type 2 diabetes mellitus. Exp Mol Med 40(5):523–532. https://doi.org/10.3858/emm.2008.40.5.523
41. Wittenbecher C, Ouni M, Kuxhaus O et al (2019) Insulin-like growth factor binding protein 2 (IGFBP-2) and the risk of developing type 2 diabetes. Diabetes 68(1):188–197. https://doi.org/10.2337/db18-0620
42. Hedbacker K, Birsoy K, Wysocki RW et al (2010) Antidiabetic effects of IGFBP2, a leptin-regulated gene. Cell Metab 11(1):11–22. https://doi.org/10.1016/j.cmet.2009.11.007
43. Wang Z, Yuan M, Xu C, Zhang Y, Ying C, Xiao X (2021) FGF21 serum levels in the early second trimester are positively correlated with the risk of subsequent gestational diabetes mellitus: a propensity-matched nested case-control study. Front Endocrinol (Lausanne) 12:630287. https://doi.org/10.3389/fendo.2021.630287
44. Xu J, Lloyd DJ, Hale C et al (2009) Fibroblast growth factor 21 reverses hepatic steatosis, increases energy expenditure, and improves insulin sensitivity in diet-induced obese mice. Diabetes 58(1):250–259. https://doi.org/10.2337/db08-0392
45. BonDurant LD, Ameka M, Naber MC et al (2017) FGF21 regulates metabolism through adipose-dependent and -independent mechanisms. Cell Metab 25(4):935–944. https://doi.org/10.1016/j.cmet.2017.03.005. (e934)
46. Wang Y, Koh WP, Yuan JM, Pan A (2017) Sex-specific association between fibroblast growth factor 21 and type 2 diabetes: a nested

case-control study in Singapore Chinese men and women. Nutr Metab (Lond) 14:63. https://doi.org/10.1186/s12986-017-0216-0
47. Reinehr T, Woelfle J, Wunsch R, Roth CL (2012) Fibroblast growth factor 21 (FGF-21) and its relation to obesity, metabolic syndrome, and nonalcoholic fatty liver in children: a longitudinal analysis. J Clin Endocrinol Metab 97(6):2143–2150. https://doi.org/10.1210/jc.2012-1221
48. Rull A, García R, Fernández-Sender L et al (2012) Serum paraoxonase-3 concentration is associated with insulin sensitivity in peripheral artery disease and with inflammation in coronary artery disease. Atherosclerosis 220(2):545–551. https://doi.org/10.1016/j.atherosclerosis.2011.11.021

Authors and Affiliations

Hong Luo[1,2] · Alina Bauer[1] · Jana Nano[1,2] · Agnese Petrera[3] · Wolfgang Rathmann[4,5] · Christian Herder[5,6,7] · Stefanie M. Hauck[3,8] · Benjamin B. Sun[9] · Annika Hoyer[10] · Annette Peters[1,2,8] · Barbara Thorand[1,8]

1 Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH), Neuherberg, Germany

2 Institute for Medical Information Processing, Biometry and Epidemiology (IBE), Faculty of Medicine, LMU Munich, Pettenkofer School of Public Health, Munich, Germany

3 Metabolomics and Proteomics Core, Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH), Neuherberg, Germany

4 Institute for Biometrics and Epidemiology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine Universität Düsseldorf, Düsseldorf, Germany

5 German Center for Diabetes Research (DZD), Partner Düsseldorf, Neuherberg, Germany

6 Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine Universität Düsseldorf, Düsseldorf, Germany

7 Department of Endocrinology and Diabetology, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine Universität Düsseldorf, Düsseldorf, Germany

8 German Center for Diabetes Research (DZD), Partner München-Neuherberg, Neuherberg, Germany

9 Translation Sciences, Research & Development, Biogen Inc., Cambridge, MA, USA

10 Biostatistics and Medical Biometry, Medical School OWL, Bielefeld University, Bielefeld, Germany