

SCIENTIFIC REPORTS



OPEN

Using a Targeted Proteomics Chip to Explore Pathophysiological Pathways for Incident Diabetes—The Malmö Preventive Project

John Molvin^{1,2}, Manan Pareek^{3,4}, Amra Jujic¹, Olle Melander^{1,5}, Lennart Råstam¹, Ulf Lindblad⁶, Bledar Daka⁶, Margrét Leósdóttir^{1,2}, Peter M. Nilsson^{1,5}, Michael H. Olsen^{3,7} & Martin Magnusson^{1,2,8}

Multiplex proteomic platforms provide excellent tools for investigating associations between multiple proteins and disease (e.g., diabetes) with possible prognostic, diagnostic, and therapeutic implications. In this study our aim was to explore novel pathophysiological pathways by examining 92 proteins and their association with incident diabetes in a population-based cohort (146 cases of diabetes versus 880 controls) followed over 8 years. After adjusting for traditional risk factors, we identified seven proteins associated with incident diabetes. Four proteins (*Scavenger receptor cysteine rich type 1 protein M130*, *Fatty acid binding protein 4*, *Plasminogen activator inhibitor 1* and *Insulin-like growth factor-binding protein 2*) with a previously established association with incident diabetes and 3 proteins (*Cathepsin D*, *Galectin-4*, *Paraoxonase type 3*) with a novel association with incident diabetes. *Galectin-4*, with an increased risk of diabetes, and *Paraoxonase type 3*, with a decreased risk of diabetes, remained significantly associated with incident diabetes after adjusting for plasma glucose, implying a glucose independent association with diabetes.

The worldwide prevalence of type 2 diabetes has risen steadily from 108 million in 1980 to 422 million in 2014 and constitutes a major threat to public health through increased morbidity and mortality¹. Almost half of all deaths attributable to hyperglycemia occur before the age of 70 years, highlighting the need for early identification and lifestyle interventions of high-risk individuals as well as identifying novel therapeutic targets². Although often simplified to a combination of insulin resistance and insulin deficiency, much remains to be explored regarding the complex pathogenic processes underlying the disease. This creates a rationale for applying a multi-system approach, including the exploration of pathophysiological pathways that may have diagnostic, prognostic, and therapeutic implications.

The recently developed proximity extension assay technology³ has enabled simultaneous analyses of large sets of proteins in small biological sample volumes. We used such an immunoassay designed to analyze 92 proteins with proposed involvement in inflammation / immunity, cardiovascular disease, and metabolism, in order to explore potential pathophysiological pathways for incident diabetes in a population-based cohort.

¹Department of Clinical Sciences, Lund University, Clinical Research Center, Malmö, Sweden. ²Department of Cardiology, Skåne University Hospital Malmö, Malmö, Sweden. ³Cardiology Section, Department of Internal Medicine, Holbæk Hospital, Holbæk, Denmark. ⁴Brigham and Women's Hospital Heart & Vascular Center, Harvard Medical School, Boston, Massachusetts, USA. ⁵Department of Internal Medicine, Skåne University Hospital, Malmö, Sweden. ⁶Institute of Medicine, Department of Public Health and Community Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. ⁷Centre for Individualized Medicine in Arterial Diseases (CIMA), Odense University Hospital, University of Southern Denmark, Odense, Denmark. ⁸Wallenberg Center for Molecular Medicine, Lund University, Lund, Sweden. Correspondence and requests for materials should be addressed to J.M. (email: johnmolvin@gmail.com)

Materials and Methods

Study sample. During 1974–1992, specific birth cohorts between 1921 and 1949 of inhabitants in Malmö, Sweden, were invited to participate in a large cohort study, i.e., the *Malmö Preventive Project (MPP)*, with a total of 33,346 individuals attending (attendance rate 71%). Re-examination of 18,238 *MPP* survivors, who were still residing in the Malmö area, the *MPP Re-Examination Study (MPP-RES)*, was conducted during 2002–2006 (attendance rate 72%). In a subsample of 1,792 participants, echocardiography was performed. These subjects were randomly selected from groups defined by glucometabolic status: normal fasting glucose, impaired fasting glucose, new onset diabetes and prevalent diabetes, with oversampling in the groups with glucometabolic disturbances to ensure numerical balance, as described previously⁴. The reason for this oversampling was to ensure sufficient numbers in each group as the study originally was designed to investigate myocardial structure and function in elderly subjects in relation to their glucometabolism. Data on lifestyle and medical history were obtained through a self-administered questionnaire. Physical activity was self-reported and categorized into 4 levels from sedentary lifestyle to physically active at a great extent. Height and weight were measured and body mass index (BMI, kg/m²) subsequently calculated. Blood pressure was measured twice in the supine position after 10 minutes of rest, and blood samples were drawn after an overnight fast and stored at –80 °C. Hypertension was defined as systolic blood pressure (SBP) >140 mmHg and diastolic blood pressure (DBP) >90 mmHg or the use of anti-hypertensive medication. Plasma samples from a total of 1,737 individuals from this subsample were successfully analyzed with the Olink Proseek Multiplex CVD III 96 × 96 proximity extension assay. Patients with missing covariates at baseline (n = 30) and prevalent diabetes (n = 681) were excluded, resulting in 1026 eligible subjects for the main analyses of incident diabetes.

All participants signed a written informed consent form before entering *MPP-RES*. The study was approved by The Regional Ethical Review board at Lund University, Sweden (LU 244-02) and complied with the Helsinki Declaration.

Proteomic Profiling. Plasma levels of proteins were analyzed by the Proximity Extension Assay (PEA) technique using the Proseek Multiplex CVD III 96 × 96 reagents kit (Olink Bioscience, Uppsala, Sweden) which uses two oligonucleotide-labeled highly specific antibodies to bind to each target protein, which allows the formation of a polymerase chain reaction sequence that can then be detected and quantified³. The CVD III panel, published in several well-renowned journals^{5–7}, consists of ninety-two proteins, carefully selected by leading experts in the field, with either established or proposed association with cardiovascular disease, inflammation and metabolism. The CVD I panel partially overlapping with the CVD III panel has previously been used to explore potential biomarkers for insulin resistance⁸ but no similar studies have been performed using the CVD III panel. The CVD III panel was also recently used for replication in a publication describing the genomic atlas of the human plasma proteome⁹. All data are presented as arbitrary units. One protein was below detectable limits in >15% samples (N-terminal pro-brain natriuretic peptide (Nt-proBNP)). Across all 92 assays, the mean intra-assay and inter-assay variations were observed to be 8.1% and 11.4%, respectively. Validation data and coefficients of variance for all proteins can be found in the online supplemental material (Validation data CVD III) and further technical information about the assays are available on the Olink homepage (<http://www.olink.com>).

Laboratory Assays. All fasting analyses (plasma glucose, serum high-density lipoprotein (HDL), and serum triglycerides (TG)) were performed at the Department of Clinical Chemistry, Malmö University Hospital, attached to a national standardization and quality control system (Beckman Coulter LX20, Beckman Coulter Inc., Brea, USA). Plasma cystatin C was analysed with an automated particle-enhanced immunoturbidimetric method, using reagents from DakoCytomation (Glostrup, Denmark).

Classification of prevalent and incident diabetes in MPP-RES. Prevalent diabetes at baseline was defined as a self-reported physician diagnosis of diabetes, use of antidiabetic medication, a diagnosis of diabetes in any of the local or national diabetes registries prior to study entry, or two separate fasting plasma glucose measurements of ≥7.0 mmol/L when available. Incident diabetes was retrieved through record linkage of the Swedish personal identification number with national and regional registries as follows: The Malmö HbA_{1c} Register that analyzed all HbA_{1c} samples at the Department of Clinical Chemistry obtained in institutional and non-institutional care in Malmö from 1988 and onwards¹⁰; The Swedish National Diabetes Register¹¹; The Regional Diabetes 2000 Register of the Skåne Region¹²; The Swedish National Patient Register covering all somatic and psychiatric hospital discharges and hospital based outpatient care¹³; The Swedish Cause-of-Death Register¹⁴; and The Swedish Prescribed Drug Register (prescription of anti-diabetic medication)¹⁵. Type of diabetes was not specified from all registries but given the mean age of the study population and since all prevalent cases of diabetes were excluded, it is reasonable to assume that an absolutely overwhelming majority of the incident cases of diabetes were type 2 diabetes.

Statistical Analysis. Non-normally distributed variables (all 91 proteins, TG, HDL, glucose and cystatin C) were ln-transformed prior to analysis. Cox proportional-hazards regression models and Harrell's concordance index (C-index)¹⁶ were used to calculate hazard ratios (HRs) for incident diabetes per standard deviation (SD) of change of log-transformed values in age- and sex-adjusted models (model 1). Proportional hazard assumption was tested using Schoenfeld residuals. Only proteins that remained significant after Bonferroni correction ($0.05/91 = 5.5 \times 10^{-4}$) in model 1 were further tested in the multivariable Cox regression model and Harrell's C-index (model 2), which was adjusted for age, sex, BMI, hypertension, antihypertensive treatment, TG, HDL, cystatin C and physical activity and furthermore in model 3 (entering fasting plasma glucose at baseline on top of model 2). The proteins associated with incident diabetes in model 1 were also tested for association with prevalent diabetes with binary logistic regression in models 1, 2 and 3.

	All subjects (n = 1707)	Subjects without prevalent diabetes (n = 1026)	Subjects with prevalent diabetes (n = 681)	p-value
Age (years)	67.4 (±6.0)	66.9 (±6.1)	68.1 (±5.9)	<0.001
Sex (% female)	498 (29.1)	331 (32.2)	167 (24.4)	0.001
BMI	28.3 (±4.3)	27.4 (±3.9)	29.8 (±4.6)	<0.001
SBP (mmHg)	146.9 (±19.8)	145.3 (±19.2)	149.1 (±20.6)	<0.001
HT (%)	1069 (62.6)	570 (55.6)	499 (73.3)	<0.001
FPG (mmol/l)	6.2 (5.6–7.4)	5.8 (5.3–6.2)	7.8 (7.1–9.2)	<0.001
TG (mmol/l)	1.3 (0.9–1.8)	1.1 (0.8–1.6)	1.5 (1.0–2.0)	<0.001
HDL (mmol/l)	1.3 (1.0–1.5)	1.3 (1.1–1.6)	1.2 (1.0–1.4)	<0.001
Cystatin C (mg/l)	1.06 (0.95–1.20)	1.05 (0.95–1.19)	1.08 (0.95–1.24)	0.002

Table 1. Baseline Characteristics of Study Participants with and without Prevalent Diabetes at Baseline examination. BMI; body mass index, SBP; systolic blood pressure; DBP; diastolic blood pressure, HT; hypertension, FPG; fasting plasma glucose, TG; triglycerides, HDL; high-density lipoprotein cholesterol. Values are displayed as means (±standard deviation) or, for skewed variables, medians and interquartile (25–75) range.

	All subjects (n = 1026)	Subjects without incident diabetes (n = 880)	Subjects with incident diabetes (n = 146)	p-value
Age (years)	66.9 (±6.1)	66.9 (±6.1)	66.7 (±5.8)	0.071
Sex (% female)	331 (32.2)	290 (32.8)	41 (28.1)	0.024
BMI	27.4 (±3.9)	27.0 (±3.7)	29.4 (±4.1)	<0.001
SBP (mmHg)	145.3 (±19.2)	145.3 (±19.5)	145.6 (±17.0)	0.002
HT (%)	571 (55.6)	467 (53.0)	104 (71.2)	0.002
FPG (mmol/l)	5.8 (5.3–6.2)	5.7 (5.3–6.2)	6.3 (6.1–6.6)	<0.001
TG (mmol/l)	1.1 (0.8–1.6)	1.1 (0.8–1.5)	1.3 (1.0–1.7)	<0.001
HDL (mmol/l)	1.3 (1.1–1.6)	1.4 (1.1–1.6)	1.2 (1.0–1.4)	<0.001
Cystatin C (mg/l)	1.05 (0.95–1.19)	1.01 (0.95–1.18)	1.08 (0.95–1.22)	0.114

Table 2. Baseline Characteristics of Study Participants with and without Incident Diabetes at Baseline Examination. BMI; body mass index, SBP; systolic blood pressure; FPG; fasting plasma glucose, HT; hypertension, TG; triglycerides, HDL; high-density lipoprotein cholesterol. Values are displayed as means (±standard deviation) or, for skewed variables, medians and interquartile (25–75) range.

All analyses were performed using SPSS Statistics version 22.0 (IBM, Armonk, New York, USA).

Results

Baseline characteristics of subjects with (n = 681) and without (n = 1026) prevalent diabetes are listed in Table 1. Subjects with prevalent diabetes at baseline had higher TG and lower HDL levels, higher BMI, increased prevalence of hypertension, and worse renal function as measured by cystatin C (Table 1). Baseline characteristics of the 1026 subjects examined for incident diabetes are listed in Table 2. Of these, 146 developed diabetes during the median follow-up time of 8.0 years (interquartile range 12 years). Subjects with incident diabetes were more often male, had higher blood pressure, TG and lower HDL levels, as well as greater BMI at baseline, compared with those who did not develop diabetes.

Associations of proteins with incident diabetes. In age- and sex-adjusted Cox analyses (model 1), 7 proteins were associated with incident diabetes and fulfilled the prespecified Bonferroni-corrected p-value of $<5.5 \times 10^{-4}$: paraoxonase-3 (PON3) ($p = 3.3 \times 10^{-9}$), fatty acid binding protein -4, (FABP4) ($p = 9.3 \times 10^{-9}$), plasminogen activator inhibitor 1 (PAI) ($p = 4.0 \times 10^{-8}$), insulin-like growth factor-binding protein 2 (IGFBP-2) ($p = 2.9 \times 10^{-7}$), scavenger receptor cysteine rich type 1 protein M130 (CD163) ($p = 3.9 \times 10^{-6}$), cathepsin D (CTSD) ($p = 5.2 \times 10^{-4}$) and Galectin-4 (Gal-4) ($p = 5.4 \times 10^{-4}$). (Table 3). Age- and sex adjusted Cox regression analysis examining all 91 proteins association to incident diabetes can be found in Supplemental Table 1.

When further adjusting for established risk factors (model 2), all 7 proteins remained significantly associated with incident diabetes; 5 proteins (CD163, Gal-4, CTSD, PAI and FABP4) with an increased risk of diabetes and 2 proteins (PON3 and IGFBP-2) with a decreased risk for incident diabetes: (Table 3). When further entering fasting plasma glucose (highly associated with incident diabetes; HR 1.30, 95% CI: 1.25–1.35; $p = 9.1 \times 10^{-39}$) as a covariate (model 3), the following four proteins remained significantly associated with increased risk of diabetes; PAI, Gal-4, CD163 and FABP4. Only PON3 remained significantly associated with decreased risk of diabetes (Table 3).

Associations of proteins with prevalent diabetes. All 7 proteins associated with incident diabetes in model 1 were significantly associated (p -values $<5.5 \times 10^{-4}$) with prevalent diabetes in a binary logistic regression model 1. However, in the fully adjusted model 3 only Gal-4 and PAI were nominally significantly associated with prevalent diabetes. (Table 4)

	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value
	Model 1		Model 2		Model 3	
PON 3	0.65 (0.56–0.75)	3.3×10^{-9}	0.79 (0.67–0.93)	0.005	0.81 (0.69–0.96)	0.014
FABP4	1.74 (1.44–2.10)	9.3×10^{-9}	1.46 (1.16–1.84)	0.001	1.48 (1.17–1.87)	0.001
PAI	1.70 (1.41–2.05)	4.0×10^{-8}	1.50 (1.21–1.84)	<0.0001	1.40 (1.14–1.72)	0.001
IGFBP-2	0.66 (0.56–0.77)	2.9×10^{-7}	0.82 (0.68–0.99)	0.039	0.89 (0.74–1.08)	0.24
CD163	1.50 (1.26–1.77)	3.0×10^{-6}	1.34 (1.11–1.60)	0.002	1.22 (1.02–1.46)	0.029
CTSD	1.33 (1.13–1.56)	5.2×10^{-4}	1.20 (1.00–1.43)	0.050	1.08 (0.91–1.29)	0.39
Gal-4	1.37 (1.15–1.64)	5.4×10^{-4}	1.30 (1.08–1.56)	0.005	1.27 (1.07–1.52)	0.008

Table 3. Cox Regression Analysis Examining Proteins relation to Incident Diabetes. Cox regressions for incident diabetes (146 cases vs. 880 controls in the MPP) adjusted for age and sex (Model 1) and age, sex, BMI, HTN, lnTG, lnHDL, Incystatin C and physical activity (Model 2) and age, sex, BMI, HTN, lnTG, lnHDL, Incystatin C, physical activity and lnglucose (Model 3). PON 3; Paraoxonase, FABP4; fatty acid binding protein 4, PAI; Plasminogen activator inhibitor 1, IGFBP-2; Insulin-like growth factor-binding protein 2, CD163; scavenger receptor cysteine rich type 1 protein M130, CTSD; Cathepsin D, Gal-4; Galectin-4.

	OR (95% CI)	p-value	OR (95% CI)	p-value	OR (95% CI)	p-value
	Model 1		Model 2		Model 3	
PAI	1.35 (1.22–1.50)	2.0×10^{-8}	1.00 (0.89–1.13)	0.984	0.80 (0.67–0.95)	0.012
FABP4	1.61 (1.44–1.81)	3.6×10^{-16}	1.16 (1.01–1.34)	0.049	1.02 (0.82–1.27)	0.847
CD163	1.58 (1.40–1.78)	8.3×10^{-14}	1.35 (1.20–1.52)	3.5×10^{-7}	1.17 (0.99–1.38)	0.061
Gal4	1.97 (1.76–2.22)	6.8×10^{-30}	1.85 (1.63–2.10)	1.7×10^{-21}	1.54 (1.29–1.84)	1.0×10^{-6}
PON3	0.59 (0.53–0.66)	1.7×10^{-21}	0.78 (0.45–1.35)	6.6×10^{-5}	0.94 (0.79–1.12)	0.495
IGFBP2	0.60 (0.54–0.67)	4.2×10^{-20}	0.77 (0.68–0.88)	1.1×10^{-4}	1.04 (0.86–1.26)	0.672
CTSD	1.73 (1.55–1.92)	1.4×10^{-23}	1.46 (1.30–1.65)	2.3×10^{-10}	1.14 (0.96–1.35)	0.140

Table 4. Logistic Regression Analysis Examining Proteins relation to Prevalent Diabetes Logistic regressions for prevalent diabetes (681 cases in the MPP) adjusted for age and sex (Model 1) and age, sex, BMI, SBP, HT, TG, HDL, physical activity and cystatin C (Model 2) and age, sex, BMI, SBP, HT, TG, HDL, physical activity, cystatin C and FPG (Model 3). PAI; Plasminogen activator inhibitor 1, FABP4; fatty acid binding protein, CD163; scavenger receptor cysteine rich type 1 protein M130, Gal-4; Galectin-4, PON 3; Paraoxonase, IGFBP-2; Insulin-like growth factor-binding protein 2, CTSD; Cathepsin D.

Harrell's concordance index models. The basic model 1 yielded a C-index of 0.542 and an addition of any one of the 7 proteins resulted in a gain in C-statistics ranged from 4.4–10.7 percentage-units. Furthermore, an addition any one of the 7 proteins to the basic model 2 (C-index 0.692) resulted in a gain in C-statistic ranged from 0.02–1.4 percentage-units.

Finally, as compared with the basic model 3 (C-index 0.780) additions any one of the 5 proteins resulted in a gain in C-statistic that ranged from 0–0.5 percentage-units (Supplementary Table 2).

Discussion

In this community-based sample of 1026 older individuals without known diabetes, we identified 7 proteins associated with incident diabetes. To the best of our knowledge, 3 of these associations (CTSD, Gal-4 and PON3) have not been previously reported.

Proteins with a previously established association with incident diabetes. *Scavenger receptor cysteine rich type 1 protein M130 (CD163).* Our findings are in line with a large prospective cohort study, which found a significantly increased risk of incident diabetes in subjects with high baseline CD163 levels¹⁷. CD163 is implicated in adipose tissue inflammation and may represent a glucose-independent mechanism in diabetes¹⁷.

Fatty acid binding protein, adipocyte (FABP4). Increased FABP4 has earlier been associated with diabetes¹⁸. FABP4 may act as a mediator between diabetes and obesity due to its role in lipid metabolism and glucose utilization¹⁹.

Plasminogen activator inhibitor 1 (PAI-1). A recent meta-analysis supported a link between PAI-1 and incident diabetes²⁰, which is in concert with our findings that also imply the association to be glucose-independent. In addition, alleles of various single nucleotide polymorphisms (SNPs) which elevate plasma PAI-1, are individually associated with type 2 diabetes²¹, suggesting a causal relationship.

Insulin-like growth factor-binding protein 2 (IGFBP-2). Inter-individual heterogeneity in endogenous IGFBP levels may influence the risk of developing type 2 diabetes²² and in a prospective nested case-control investigation,

plasma IGFBP-2 levels were strongly and inversely associated with the risk of diabetes²³ which is consistent with the protective effects of IGFBP-2 seen in our study.

Proteins with a novel association with incident diabetes. *Cathepsin D (CTSD).* A recent proteomic study showed a cross-sectional association between CTSD and prevalent insulin resistance⁸. This finding together with our finding that CTSD is associated with both prevalent and incident diabetes suggest that CTSD may have a mechanistic role in the development of diabetes and insulin resistance. The main effects of the lysosomal endopeptidase CTSD include intracellular protein turnover and extracellular matrix breakdown²⁴. It has been suggested that CTSD acts a mediator between obesity and chronic adipose tissue inflammation as weight gain has shown to stimulate CTSD activity leading to adipocyte apoptosis, which is an important contributor to insulin resistance²⁵. Furthermore, increased CTSD activity has in experimental studies been shown to be involved in the truncation of ApoA1 (the most abundant protein in HDL) to ApoA1 Δ (1–38); a variant which is more abundant in patients with diabetes and more susceptible to oxidation²⁶.

Galectin-4 (Gal-4). Gal-4 is a small lectin protein expressed almost exclusively in the gastrointestinal tract and is involved in protein apical trafficking and lipid raft stabilization i.e. the transport of proteins from inside the cell to the cell membrane. One of the proteins transported from the Golgi apparatus to the apical cell membrane of the enterocyte is the protease dipeptidyl peptidase-4 (DPP-4)²⁷. DPP-4's most known effect is the inactivation of our two most abundant incretins; glucose-dependent insulinotropic polypeptide (GIP) and proglucagon-derived peptide glucagon-like peptide-1 (GLP-1)²⁸. GLP-1-analogues and DPP-4-inhibitors are well-established treatments in type 2 diabetes and recently two major studies of GLP-1-analogues have shown, in addition to lowering blood glucose, a reduced risk of cardiovascular disease^{29,30} and mortality³⁰. One possible explanation of our finding that Gal-4 is associated with both incident and prevalent diabetes is that increased expression of Gal-4 leads to increased activity of DPP-4 and thus reduced activity of GLP-1 and increased risk of diabetes and cardiovascular complications. Although other galectins (e.g. Gal-3³¹ and Gal-1³²) have been associated with diabetes, no association of Gal-4 with diabetes has, to our knowledge, been reported before.

Paraoxonase type 3 (PON3). PON3 is similar to paraoxonase type 1 (PON1) in activity but differs from it in substrate specificity³³. Both PON3 and PON1 are bound to HDL and because of their similar properties as antioxidants, it is possible PON3 also plays a role in the prevention of LDL and HDL oxidation³⁴. Previous studies have consistently reported that PON1 is lower in patients with diabetes compared to control subjects³⁵. Although we could not find previous data regarding PON3 in plasma and subsequent risk of diabetes, there are studies that have described lower levels of PON3 with an increased duration of diabetes and in patients with diabetes and coronary artery disease (CAD) compared to subjects with diabetes without CAD^{36,37}. All these findings are in line with the diabetes protective effects of PON3 seen in our study.

Study limitations. Since type 2 diabetes is a multifactorial disease with a range of known risk factors contributing to its pathogenesis, these risk factors should be considered when conclusions are drawn regarding associations. Although we attempted adjustment for a heterogeneous panel of risk factors, the observational nature of this study prevents us from ruling out that other confounders may have affected the outcome of our analysis. Furthermore, we did not have the possibility for repeated or confirmatory measurements of the proteins through an additional method. Baseline HbA1c was missing in >30% of the subjects and therefore excluded which is a weakness as HbA1c is a very strong predictor for incident diabetes. There was no oral glucose tolerance test performed in these subjects. Moreover, our data was collected at a single regional center, without the option of replicating the findings although we attempted to limit the risk of false positive findings by Bonferroni correction. The original selection of the population with oversampling of groups based on glucometabolic disturbances mentioned in the *Methods* section can raise concerns how well this cohort represents the background population but the rate of incidence of diabetes in this cohort is comparable to other similar cohorts^{38,39}. Furthermore, as mentioned in the *Methods* section, type of diabetes was not specified from the registries but we have assumed that the incidence of type 1 diabetes must be extremely low due to the participants' mean age of 67.4 (\pm 6.0) years at the baseline examination.

Lastly, although the CVD III panel is only partially directed towards metabolism, it also includes proteins associated with cardiovascular disease and inflammation and thus a more specifically designed assay towards diabetes and/or metabolism could possibly have revealed additional findings.

Conclusion

Our study confirmed previously established associations with incident diabetes for CD163, FABP4, PAI, and IGFBP-2. Furthermore, we identified novel associations for CTSD, Gal-4 and PON3 with incident diabetes. Gal-4 and PON3 remained significantly associated with incident diabetes after adjusting for plasma glucose, implying a glucose independent association with diabetes. None of the proteins showed a substantial increase in C-index which, at present, would not warrant clinical use as a biomarker. Nevertheless, the associations of these three proteins could represent novel biological mechanisms, broadening our understanding of the complex pathogenesis of diabetes. First and foremost, our results merit replication in an independent cohort and if successful, future prospective studies to clarify their role in the possible pathogenesis of diabetes.

References

1. Collaboration, N. C. D. R. F. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet* **387**, 1513–1530, [https://doi.org/10.1016/S0140-6736\(16\)00618-8](https://doi.org/10.1016/S0140-6736(16)00618-8) (2016).
2. Mathers, C. D. & Loncar, D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* **3**, e442, <https://doi.org/10.1371/journal.pmed.0030442> (2006).

3. Assarsson, E. *et al.* Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One* **9**, e95192, <https://doi.org/10.1371/journal.pone.0095192> (2014).
4. Leosdottir, M. *et al.* Myocardial structure and function by echocardiography in relation to glucometabolic status in elderly subjects from 2 population-based cohorts: a cross-sectional study. *Am Heart J* **159**, 414–420 e414, <https://doi.org/10.1016/j.ahj.2009.12.028> (2010).
5. Brunner, P. M. *et al.* The atopic dermatitis blood signature is characterized by increases in inflammatory and cardiovascular risk proteins. *Sci Rep* **7**, 8707, <https://doi.org/10.1038/s41598-017-09207-z> (2017).
6. Ouwerkerk, W. *et al.* Biomarker-Guided Versus Guideline-Based Treatment of Patients With Heart Failure: Results From BIOSTAT-CHF. *J Am Coll Cardiol* **71**, 386–398, <https://doi.org/10.1016/j.jacc.2017.11.041> (2018).
7. Enroth, S. *et al.* Systemic and specific effects of antihypertensive and lipid-lowering medication on plasma protein biomarkers for cardiovascular diseases. *Sci Rep* **8**, 5531, <https://doi.org/10.1038/s41598-018-23860-y> (2018).
8. Nowak, C. *et al.* Protein Biomarkers for Insulin Resistance and Type 2 Diabetes Risk in Two Large Community Cohorts. *Diabetes* **65**, 276–284, <https://doi.org/10.2337/db15-0881> (2016).
9. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79, <https://doi.org/10.1038/s41586-018-0175-2> (2018).
10. Enhorning, S. *et al.* Plasma copeptin and the risk of diabetes mellitus. *Circulation* **121**, 2102–2108 (2010).
11. Cederholm, J. *et al.* Risk prediction of cardiovascular disease in type 2 diabetes: a risk equation from the Swedish National Diabetes Register. *Diabetes care* **31**, 2038–2043, <https://doi.org/10.2337/dc08-0662> (2008).
12. Lindholm, E., Agardh, E., Tuomi, T., Groop, L. & Agardh, C. D. Classifying diabetes according to the new WHO clinical stages. *European journal of epidemiology* **17**, 983–989 (2001).
13. Ludvigsson, J. F. *et al.* External review and validation of the Swedish national inpatient register. *BMC public health* **11**, 450, <https://doi.org/10.1186/1471-2458-11-450> (2011).
14. A finger on the pulse. Monitoring public health and social conditions in sweden 1992–2002. (2003) Stockholm, National Board of Health and Welfare. (2003).
15. Wettermark, B. *et al.* The new Swedish Prescribed Drug Register—opportunities for pharmacoepidemiological research and experience from the first six months. *Pharmacoepidemiology and drug safety* **16**, 726–735, <https://doi.org/10.1002/pds.1294> (2007).
16. Harrell, F. E. *et al.* Evaluating the yield of medical tests. *JAMA: the journal of the American Medical Association* **247**, 2543–2546 (1982).
17. Moller, H. J., Frikke-Schmidt, R., Moestrup, S. K., Nordestgaard, B. G. & Tybjaerg-Hansen, A. Serum soluble CD163 predicts risk of type 2 diabetes in the general population. *Clin Chem* **57**, 291–297, <https://doi.org/10.1373/clinchem.2010.154724> (2011).
18. Djousse, L. *et al.* Plasma fatty acid-binding protein 4, nonesterified fatty acids, and incident diabetes in older adults. *Diabetes care* **35**, 1701–1707, <https://doi.org/10.2337/dc11-1690> (2012).
19. Hotamisligil, G. S. *et al.* Uncoupling of obesity from insulin resistance through a targeted mutation in aP2, the adipocyte fatty acid binding protein. *Science* **274**, 1377–1379 (1996).
20. Yarmolinsky, J. *et al.* Plasminogen activator inhibitor-1 and type 2 diabetes: a systematic review and meta-analysis of observational studies. *Sci Rep* **6**, 17714, <https://doi.org/10.1038/srep17714> (2016).
21. Huang, J. *et al.* Genome-wide association study for circulating levels of PAI-1 provides novel insights into its regulation. *Blood* **120**, 4873–4881, <https://doi.org/10.1182/blood-2012-06-436188> (2012).
22. Rajpathak, S. N. *et al.* The role of insulin-like growth factor-I and its binding proteins in glucose homeostasis and type 2 diabetes. *Diabetes Metab Res Rev* **25**, 3–12, <https://doi.org/10.1002/dmrr.919> (2009).
23. Rajpathak, S. N. *et al.* Insulin-like growth factor axis and risk of type 2 diabetes in women. *Diabetes* **61**, 2248–2254, <https://doi.org/10.2337/db11-1488> (2012).
24. Masson, O. *et al.* Pathophysiological functions of cathepsin D: Targeting its catalytic activity versus its protein binding activity? *Biochimie* **92**, 1635–1643, <https://doi.org/10.1016/j.biochi.2010.05.009> (2010).
25. Eguchi, A. & Feldstein, A. E. Lysosomal Cathepsin D contributes to cell death during adipocyte hypertrophy. *Adipocyte* **2**, 170–175, <https://doi.org/10.4161/adip.24144> (2013).
26. Cubedo, J. *et al.* A novel truncated form of apolipoprotein A-I transported by dense LDL is increased in diabetic patients. *J Lipid Res* **56**, 1762–1773, <https://doi.org/10.1194/jlr.P057513> (2015).
27. Delacour, D. *et al.* Galectin-4 and sulfatides in apical membrane trafficking in enterocyte-like cells. *J Cell Biol* **169**, 491–501, <https://doi.org/10.1083/jcb.200407073> (2005).
28. Drucker, D. J. & Nauck, M. A. The incretin system: glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors in type 2 diabetes. *Lancet* **368**, 1696–1705, [https://doi.org/10.1016/S0140-6736\(06\)69705-5](https://doi.org/10.1016/S0140-6736(06)69705-5) (2006).
29. Marso, S. P. *et al.* Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes. *N Engl J Med* **375**, 1834–1844, <https://doi.org/10.1056/NEJMoa1607141> (2016).
30. Marso, S. P. *et al.* Liraglutide and Cardiovascular Outcomes in Type 2 Diabetes. *N Engl J Med* **375**, 311–322, <https://doi.org/10.1056/NEJMoa1603827> (2016).
31. Yilmaz, H., Cakmak, M., Inan, O., Darcin, T. & Akcay, A. Increased levels of galectin-3 were associated with prediabetes and diabetes: new risk factor? *J Endocrinol Invest* **38**, 527–533, <https://doi.org/10.1007/s40618-014-0222-2> (2015).
32. Perone, M. J. *et al.* Suppression of autoimmune diabetes by soluble galectin-1. *J Immunol* **182**, 2641–2653, <https://doi.org/10.4049/jimmunol.0800839> (2009).
33. Reddy, S. T. *et al.* Human paraoxonase-3 is an HDL-associated enzyme with biological activity similar to paraoxonase-1 protein but is not regulated by oxidized lipids. *Arteriosclerosis, thrombosis, and vascular biology* **21**, 542–547 (2001).
34. Draganov, D. I. *et al.* Human paraoxonases (PON1, PON2, and PON3) are lactonases with overlapping and distinct substrate specificities. *J Lipid Res* **46**, 1239–1247, <https://doi.org/10.1194/jlr.M400511-JLR200> (2005).
35. Triki, S. *et al.* Is PON1 a factor of cardiovascular risk in type 2 diabetes? *Ann Cardiol Angeiol (Paris)* **65**, 81–86, <https://doi.org/10.1016/j.ancard.2015.02.006> (2016).
36. Tartan, Z. *et al.* The role of paraoxonase (PON) enzyme in the extent and severity of the coronary artery disease in type-2 diabetic patients. *Heart Vessels* **22**, 158–164, <https://doi.org/10.1007/s00380-006-0957-6> (2007).
37. Shen, Y. *et al.* Association of elevated apoA-I glycation and reduced HDL-associated paraoxonase1, 3 activity, and their interaction with angiographic severity of coronary artery disease in patients with type 2 diabetes mellitus. *Cardiovasc Diabetol* **14**, 52, <https://doi.org/10.1186/s12933-015-0221-4> (2015).
38. Borne, Y. *et al.* Cadmium exposure and incidence of diabetes mellitus—results from the Malmo Diet and Cancer study. *PLoS One* **9**, e112277, <https://doi.org/10.1371/journal.pone.0112277> (2014).
39. Ebrahimi, H., Emamian, M. H., Hashemi, H. & Fotouhi, A. High Incidence of Diabetes Mellitus Among a Middle-Aged Population in Iran: A Longitudinal Study. *Can J Diabetes* **40**, 570–575, <https://doi.org/10.1016/j.cjcd.2016.05.012> (2016).

Acknowledgements

Dr. Magnusson was supported by grants from the Wallenberg Centre for Molecular Medicine, Lund University (ALFSKANE-675271), Medical Faculty of Lund University (ALFSKANE-432021) (ALFSKANE-436111), Skåne University Hospital, the Crafoord Foundation, the Ernhold Lundstroms Research Foundation, Region Skåne, the Hulda and Conrad Mossfelt Foundation, the Southwest Skåne’s Diabetes Foundation, the Kocksa foundation, the Research Funds of Region Skåne and the Swedish Heart and Lung foundation (2015-0322).

Author Contributions

J.M., M.P., O.M., L.R., U.L., B.D., M.L., P.M.N., M.H.O. and M.M. contributed to study concept and design. J.M., M.L., M.M. acquired data. J.M., M.P., A.J., M.L., M.H.O. and M.M. analysed and interpreted data. J.M., M.P., M.H.O. and M.M. drafted the manuscript. J.M., M.P., A.J., O.M., L.R., U.L., B.D., M.L., P.M.N., M.H.O., M.M. critically revised the manuscript for important intellectual content. J.M., M.P., A.J., M.L., M.H.O., M.M. contributed to statistical analysis. M.P., O.M., L.R., U.L., P.M.N., M.H.O., M.M. obtained funding. P.M.N. provided administrative, technical, or material support. J.M., M.M. and P.M.N. supervised the study. J.M., P.M.N. and M.M. are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-36512-y>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019