Protein biomarkers for insulin resistance and type 2 diabetes risk in two large community cohorts

Running title:
Protein biomarkers for insulin resistance

Christoph Nowak¹, Johan Sundström², Stefan Gustafsson¹, Vilmantas Giedraitis³, Lars Lind², Erik Ingelsson¹,⁴,⁵, Tove Fall¹

¹Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden
²Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala University, Uppsala, Sweden
³Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden
⁴Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom
⁵Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA, USA

Corresponding author:
Associate Professor Tove Fall, DVM, PhD
Department of Medical Sciences, Molecular Epidemiology
Uppsala University
Box 1115, S - 751 41 Uppsala, Sweden
Phone: +46-70-2215859
E-mail: tove.fall@medsci.uu.se

Word count: 3,932
Number of tables: 3
Number of figures: 1
ABSTRACT

Insulin resistance (IR) is a precursor of type 2 diabetes (T2D) and improved risk prediction and understanding of the pathogenesis are needed. We used a novel high-throughput 92-protein assay to identify circulating biomarkers for HOMA-IR in two cohorts of non-diabetic community residents (n = 1,367; age 73 ± 3.6 years). Adjusted linear regression identified cathepsin D and confirmed six proteins (leptin, renin, interleukin-1 receptor antagonist [IL-1ra], hepatocyte growth factor, fatty acid binding protein 4, tissue plasminogen activator [t-PA]) as IR biomarkers. Mendelian randomization analysis indicated a positive causal effect of IR on t-PA concentrations. Two biomarkers, IL-1ra (hazard ratio, HR 1.28, 95% CI 1.03-1.59) and t-PA (HR 1.30, 1.02-1.65) were associated with incident T2D, and t-PA predicted 5-year transition to hyperglycemia (odds ratio 1.30, 1.02-1.65). Additional adjustment for fasting glucose rendered both coefficients insignificant and revealed an association between renin and T2D (HR 0.79, 0.62-0.99). Lasso regression suggested a risk model including IL-1ra, t-PA, and the Framingham Offspring Study T2D score, but prediction improvement was non-significant (difference in C-index 0.02, 95% CI -0.08-0.12) over the T2D score only. In conclusion, proteomic blood profiling indicated cathepsin D as a new IR biomarker and suggested a causal effect of IR on t-PA.
Worldwide, diabetes mellitus affected over 387 million people and contributed to more than 4.9 million deaths in 2014. The prevalence of diabetes mellitus is projected to increase to 592 million by 2035 (1). Decreased sensitivity to circulating insulin (i.e. insulin resistance, IR) induces compensatory hyperinsulinemia and leads to the development of type 2 diabetes (T2D) if pancreatic β-cell capacity is insufficient to maintain glucose homeostasis (2). Insulin resistance constitutes both a precursor of and a therapeutic target in hyperglycemia, and was found to be an independent risk factor for cardiovascular disease (CVD) (3), as well as a major contributor to vascular morbidity in T2D (4). Recent advances have made large-scale -omics studies possible that have pinpointed several tentative novel biomarkers for T2D, including branched-chain amino acids (5) and circulating microRNAs (6). Yet, a 2013 systematic review (7) failed to find evidence of benefit from adding novel circulating biomarkers and genetic markers to traditional T2D risk factors. Studies on biomarkers for IR have suggested several candidates, including ghrelin (8), and retinol binding protein-4 (9). The identification of novel biologic predictors for T2D and IR is crucial for improved risk assessment and may help in understanding causal pathways beyond established genetic and lifestyle-related factors.

Recently, a new technology, the proximity extension assay (PEA) (10), has enabled the simultaneous analysis 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 and CVD to explore potential biomarkers for IR. The objectives of this study were to: 1) evaluate the association of CVD/inflammatory candidate protein biomarkers with prevalent IR in two large community cohorts without diabetes; 2)
explore causal associations between biomarkers and IR in bidirectional Mendelian Randomization (MR) analysis; 3) assess the association of IR biomarkers with 10-year incident T2D and 5-year risk of transition to worse glycemia category; as well as to estimate the predictive performance of biomarkers for future T2D compared to an established risk score.
RESEARCH DESIGN AND METHODS

Cohort Characteristics

ULSAM

In 1970, all male residents (n = 2,841) of Uppsala county, Sweden, born between 1920 and 1924 were invited to participate in the Uppsala Longitudinal Study of Adult Men (ULSAM; n = 2,322 [81.7%] enrolled) (11), which includes regular assessments every five to ten years. The baseline of the present study was set to the assessment at age 77 years (839 of 1,398 invited men [59.9%]), including recent targeted proteomic serum profiling. Diabetes was defined as fasting plasma glucose ≥ 7 mmol/l and/or glycated hemoglobin HbA1c ≥ 6.5% (48 mmol/mol) at assessment ages 77, 82, and 88, and/or use of anti-diabetic medication according to the Swedish Prescribed Drug Register ATC code A10 and/or diagnosis of T2D according to the National Patient Register. Incident events of T2D were identified up to age 88. Proteomic profiling was done in 770 samples of which eight were excluded during quality control. We excluded 156 individuals with prevalent diabetes and 66 with insufficient data for confounders, leaving 540 persons for inclusion in the present study. The regional ethics review board at Uppsala University approved the study and all participants provided written informed consent.

PIVUS

In 2001, the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study group invited an unselected sample of 70-year-old residents of Uppsala community
(1,016 of 2,025 invited persons [50.2%] were enrolled; 50% female) primarily to validate measures of endothelial function (12). Baseline assessment, including recent proteomic profiling of blood plasma, was done at age 70. Diabetes was defined as plasma glucose concentration $\geq$ 7 mmol/l at ages 70, 75, and 80, and/or use of anti-diabetic medication and/or diagnosis of T2D according to validated hospital records. Incident T2D events were identified up to age 80. Among 1,003 subjects undergoing proteomic profiling, twelve were removed during quality control. We further excluded 116 individuals with prevalent diabetes and 48 with insufficient data for confounders, leaving 827 persons to be included in the present study. The regional ethics review board at Uppsala University approved the study and all participants provided written informed consent.

**Measurement of Insulin Resistance**

The homeostasis model assessment IR index (HOMA-IR) was calculated according to the method proposed by Matthews and colleagues (13) with glucose in mmol/l and insulin in mU/l (Equation 1). Plasma insulin was measured by an enzyme-linked immunosorbant assay (ELISA) in PIVUS (Boehringer, Mannheim, Germany) and ULSAM (Mercodia, Uppsala, Sweden). Glucose concentrations were quantified by the glucose dehydrogenase method (Gluc-DH by Merck, Darmstadt, Germany) in plasma from ULSAM, and with similar methods (HemoCue, Ängelholm, Sweden) in whole blood from PIVUS (converted to plasma values by adding 11%).

**Equation 1.**

\[
HOMA-IR = \left( \frac{\text{fasting glucose}_{\text{Plasma}}}{22.5} \right) \times \left( \frac{\text{fasting insulin}_{\text{Plasma}}}{22.5} \right)
\]
Proteomic Profiling

The Olink Proseek Multiplex CVD 96x96® PEA assay (10) uses two highly specific antibodies for each protein, which allows the formation of a polymerase chain reaction (PCR) reporter sequence from attached oligonucleotide strands when both antibodies are bound to the target protein’s surface. The assay requires < 10 µL sample volume and measures 92 proteins associated with CVD or inflammation and four internal control samples. Normalized Protein Expression (NPX) values were generated from quantitative PCR quantification cycle (Cq) values, where higher Cq correspond to lower protein abundance. Cq values (log₂-scale) were corrected for technical variation by an inter-plate control and lower limits of detection (LOD) were determined through a negative control (Equation 2). The validation study of the assay, which included 90 proteins and seven samples analyzed in nine separate runs, found the mean intra-assay coefficient of variation to be 8% (range 4-13%) and the inter-assay CV to be 15% (range 11-39%).

Equation 2.

\[
NPX = \text{Olink negative control} - (\Delta Cq_{\text{Sample}} - \Delta \text{Interplate control})
\]

Values below LOD were imputed as LOD / 2, normalized for plate and storage time (based on the observed and predicted values obtained from a spline model) and rescaled to a distribution with a mean of 0 and a standard deviation of 1. Quality control included removal of proteins with > 15% samples below the LOD and subjects with tail-distribution (i.e. outlying) missingness as judged by histogram (> 5% missing in PIVUS, and > 2% missing in ULSAM) were excluded. The final data set included 80 proteins. Proteins excluded from the statistical analyses were interleukin 4, melusin, natriuretic
peptide B, beta-nerve growth factor, SIR2-like protein, NF-kappa-B essential modulator, pentraxin-related protein 3, N-terminal pro-brain natriuretic peptide, matrix metalloproteinase 7, membrane-bound aminopeptidase P, heat shock 27 kDa proteins, and cathepsin B.

**Genetic Data**

For MR analyses, we used the non-weighted genetic IR score composed of ten single nucleotide polymorphisms (SNPs) validated in up to 18,565 subjects by Scott and colleagues (14) as an instrumental variable (IV) for HOMA-IR (Online Supplemental Table 1). We further identified suitable IVs for three biomarkers from a literature search (15-17). We then performed a genome-wide association study (GWAS) using the software SNPTESTv2.4 in the PIVUS and ULSAM studies for the remaining four biomarkers based on genotyping with the Illumina OmniExpress/Omni2.5 array combined with the Illumina Cardio-MetaboChip array, which was further imputed up to the 1000G March 2012 release using IMPUTE2 (18). The association of biomarker IVs with lnHOMA-IR was tested in the MAGIC (Meta-Analyses of Glucose and Insulin-related traits Consortium) cohort using publically available data (19) and for the genetic IR score with biomarkers, we used PIVUS and ULSAM. For SNPs not reported in the MAGIC data, we selected a proxy variant in strong linkage disequilibrium ($r^2 > 0.8$) via SNAP (http://www.broadinstitute.org/mpg/snap/ldsearch.php) and ascertained allele alignment with reference to the International HapMap Project CEU reference population (hapmap.ncbi.nlm.nih.gov/).
Statistical Analysis

Association of Biomarkers with Insulin Resistance and Incident Diabetes Traits

All statistical analyses were carried out in R version 3.1.1. Preliminary models indicated non-normal distribution of model residuals, and C-reactive protein (CRP) concentrations and HOMA-IR were transformed to natural logarithmic scale to alleviate non-normality. Separate linear regression models were assessed for each biomarker with lnHOMA-IR as the dependent variable and body mass index (BMI), waist circumference, lnCRP, age, comorbidity, storage time, and sex as independent variables. Comorbidity was dummy-coded based on a Charlson Comorbidity Index (20) of 0 or ≥ 1. The choice of the independent variables was based on a hypothetical causal diagram assisted by the DAGitty version 2.2 software (www.dagitty.net, Supplemental Figure 1) (21). The analysis was done in PIVUS as discovery sample and biomarkers with significant associations at a 5% false discovery rate (FDR) (22) were assessed in ULSAM for validation, also at a 5% FDR threshold. For all biomarkers, model assumptions of homoscedasticity and normality, and the impact of potential outliers were examined in plots of residuals against normal quantiles (QQ-plot), fitted values, and leverage, respectively.

Thereafter, we assessed biomarkers related to IR for associations with 10-year incident T2D using Cox regression in ULSAM and PIVUS combined, with adjustment for BMI, waist, lnCRP, age, comorbidity, storage time, cohort, and sex. We additionally included fasting glucose levels in separate models. As the date of incident diabetes, the first recorded event fulfilling the definition of diabetes as specified above was used (i.e. date
of diagnosis, anti-diabetic medication prescription, or blood glucose thresholds at five and ten-year follow ups). Individuals without an event were censored at the last assessment date or at date of death. The proportional hazards assumption was assessed using scaled Schoenfeld residual plots with formal significance testing for neutral slopes. Logistic regression analysis was used to predict five-year risk of worsening fasting glycemia category (< 5.6 mmol/l, 5.6–6.9 mmol/l, and ≥ 7 mmol/l, or established T2D) in ULSAM from ages 77 to 82 years, and in PIVUS from ages 70 to 75 years.

**Predictive Performance and Comparison to Established Risk Factors**

To assess predictive performance of biomarkers for 10-year T2D risk, we randomly split the combined cohorts of participants with sufficient data into a two-thirds-learning (n = 911) and one-third-internal validation sample (n = 456). The Framingham Offspring Study risk score for T2D (FORS, composed of sex, BMI, age, family history of diabetes, blood pressure, high density lipoprotein cholesterol, triglycerides, and fasting glucose) (23) was calculated for each individual and used as a baseline model to assess the incremental improvement of adding biomarkers. Predictor selection was carried out in LASSO penalized Cox regression with 10-fold internal cross-validation in the learning sample and implemented with the `glmnet` package in R. In storage time and cohort-adjusted models forced to include the FORS score, we allowed predictor choice among the validated IR biomarkers and used lambda minimum to select the optimum model, which was then evaluated in the validation sample. We assessed discrimination (future case and non-case differentiation) with the receiver operating characteristic curve-based C-index (24) and compared models via likelihood ratio test. Calibration (the consistency
between observed and predicted risks) was assessed by Grønnesby-Borgan test according to May & Hosmer (25). This test is based on grouping subjects according to their risk estimates and comparing the sum of Cox model martingale residuals between groups, which assumes zero under the null hypothesis of perfect agreement between predicted and observed risks.

**Instrumental Variable Analysis**

Mendelian randomization techniques based on IV analysis were used to assess potential causal associations between biomarkers and IR in both directions (26).

**The Causal Effect of Insulin Resistance on Protein Concentrations**

We evaluated the association of the genetic IR score with lnHOMA-IR using the summary statistics for 46,186 non-diabetic individuals in the MAGIC cohort based on the method described by Dastani and colleagues (27) and implemented via `gtx` in R. The association of the genetic IR score with biomarkers was assessed in PIVUS and ULSAM separately in age- and sex-adjusted linear regression models and meta-analyzed using a fixed effect, standard error-weighted model via `metafor` in R (Online Supplemental Table 2). The IV estimator $\beta_{IV}$ was calculated as the ratio of two regression coefficients based on the Wald ratio ($\beta_{SNP-BIOMARKER} / \beta_{SNP-HOMA-IR}$). Standard errors were calculated using the delta method, which we previously validated for use in a similar setting (28), as $\text{abs}(\beta_{IV})(\text{se}_{\beta_{SNP-INTERMEDIATE}} / \beta_{SNP-INTERMEDIATE})^2 + (\text{se}_{\beta_{SNP-OUTCOME}} / \beta_{SNP-OUTCOME})^2)^{0.5}$. Causal estimators were tested at a nominal significance threshold of $P < 0.05$. Sensitivity
analysis to exclude pleiotropy of the IV was performed by comparing IV estimates for individual SNPs in forest plots (Online Supplemental Figure 2).

**The Causal Effect of Protein Concentrations on Insulin Resistance**

We assessed the association of each biomarker IV with lnHOMA-IR in the MAGIC cohort (Online Supplemental Table 3). The association of each biomarker IV with biomarker concentration was derived from either published GWAS or ULSAM and PIVUS. The IV estimator $\beta_{IV}$ was calculated as $\beta_{SNP-HOMA-IR} / \beta_{SNP-BIOMARKER}$ and standard errors were calculated using the delta method.
RESULTS

Biomarkers for Insulin Resistance

Table 1 shows baseline cohort characteristics. The design of the study is visualized in Fig. 1. We found 32 biomarkers associated with lnHOMA-IR in the PIVUS study (n = 827, 48.9% male, mean age 70.2 ± 0.2 years). Seven of these 32 biomarkers were replicated at a 5% FDR level in a sample of 540 men from the ULSAM cohort (mean age 77.6 ± 0.8 years). All subjects provided complete data for confounders. The seven identified biomarkers (leptin, tissue plasminogen activator [t-PA], renin, interleukin-1 receptor antagonist [IL-1ra], hepatocyte growth factor [HGF], cathepsin D, and fatty acid binding protein 4 [FABP-4]) were all positively associated with IR (Table 2). Of these, leptin showed the strongest association (β 0.27, 95% CI 0.22-0.33 in PIVUS; β 0.10, 95% CI 0.07-0.13 in ULSAM, where the coefficients represent the change in lnHOMA-IR associated with an SD-unit increase in NPX value).

Bivariate Pearson correlations indicated positive associations (P < 0.05) between all seven biomarkers except for leptin and renin (r = 0.05, P = 0.058) in the low to moderate range (0.15-0.69).

Association with Incident Diabetes and Worsening Hyperglycemia

During follow-up (mean 9.7 ± 0.5 years), there were 73 and 38 incident cases of T2D in PIVUS and ULSAM, respectively, among 1,367 participants. In separate models adjusted for cohort and confounders, two biomarkers were associated with increased T2D risk
(Table 3): IL-1ra (HR 1.28, 95% CI 1.03-1.59) and t-PA (HR 1.30, 95% CI 1.03-1.65).
Additional adjustment for fasting glucose rendered both associations non-significant (t-PA HR 1.14, 95% CI 0.91-1.44; IL-1ra HR 1.19, 95% CI 0.94-1.50) and revealed a negative association with renin levels (HR 0.79, 95% CI 0.62-0.99).

At five-year follow-up assessment, there were 115 and 88 cases of worse glycemic state compared to baseline in PIVUS and ULSAM, respectively. In adjusted logistic regression analysis, increased concentrations of t-PA (OR 1.23, 95% CI 1.02-1.48) predicted worse glycemic status at five-year follow-up (Table 3).

Comparison to Established Risk Factors for Type 2 Diabetes

In LASSO Cox regression based on the learning sample, 10-year diabetes risk was predicted by a model that included the FORS score and the two biomarkers associated with T2D (t-PA and IL-1ra). In the internal validation set using the proposed beta coefficients, this new model improved the C-index compared to the FORS score-only model by 0.022, from C = 0.801 (95% CI 0.701-0.991) to C = 0.823 (95% CI 0.723-0.923). There was no significant difference in model fit ($\chi^2 = 5.258, P = 0.07$), with the biomarker model explaining 12.7% of the variance compared to 11.4% by the FORS score-only model. Both models demonstrated adequate calibration; Grønnesby-Borgan $X^2 = 5.378, P = 0.25$ for baseline, and $X^2 = 2.399, P = 0.66$ for the full model, respectively.
Causal Associations between Insulin Resistance and Biomarkers

In a literature search, we identified suitable IVs for IL-1ra (rs4251961 and rs6759676) (15), HGF (rs5745687) (16), and t-PA (rs9399599, rs3136739, and rs7301826) (17). In a GWAS for the remaining four biomarkers using the conventional threshold for genome-wide significance ($P < 5 \times 10^{-8}$), we found a suitable genetic IV for cathepsin D only (rs55861089; Online Supplemental Table 3).

We found evidence of a causal effect of HOMA-IR on t-PA concentrations ($\beta_{IV} 3.21$, 95% CI 0.72-5.70, $P = 0.012$). No evidence of a causal effect of IR on any of the other biomarkers was found (Online Supplemental Table 2). Sensitivity analysis for the genetic instrument did not indicate pleiotropic effects (Online Supplemental Figure 2).

For the causal effect of biomarker on HOMA-IR, we identified suitable genetic IVs for IL-1ra, t-PA, HGF, and cathepsin D. The results of IV analysis in MAGIC did not show evidence of a causal effect of any of these biomarkers on IR (Online Supplemental Table 3).
DISCUSSION

In two prospective community samples of 1,367 elderly non-diabetic individuals, we identified seven proteins positively associated with prevalent IR, and one of these, Cathepsin D, has not previously been reported as associated with IR. The correlations between the protein concentrations were weak to moderate. These correlations could either be explained by biomarkers being implicated in the same biological pathways, or they could represent different biological cascades related to IR and cardiovascular risk. We therefore carried all seven biomarkers forward for further analysis, where we found evidence for a causal effect of IR on t-PA concentrations suggesting an effect of IR on blood coagulation and extracellular matrix modeling - important components of atherosclerosis. We also found higher baseline concentrations of t-PA and IL-1ra were associated with ten-year diabetes risk, and t-PA predicted worse five-year fasting glucose levels. Compared to an established diabetes risk score, the addition of biomarkers did not improve discrimination significantly.

Causal Effect of Insulin Resistance on Tissue Plasminogen Activator

Mendelian randomization analysis offers a statistical approach to inferring causality in observational studies. As variants of genetic alleles are randomly inherited at conception, their distributions are free from confounding influences and reverse causation. In MR analysis, a genetic variant or combination of variants known to be associated with an intermediate phenotype is used as the instrumental variable to assess the possible causal effect of the intermediate on the outcome variable (26).
Using MR analysis, we found evidence for a positive causal effect of IR on t-PA antigen levels, which has not previously been reported although their correlation is well established (29). Tissue plasminogen activator is expressed by endothelial cells and acts mainly by converting plasminogen to plasmin, thus contributing to fibrinolysis and extracellular matrix remodeling. It also acts as a pro-inflammatory cytokine (30). Circulating t-PA activity is regulated through complex formation with its main inhibitor plasminogen activator inhibitor 1 (PAI-1), a major source of which is adipose tissue (31). Elevated PAI-1 activity and concentration are both associated with raised t-PA levels (32) and the observed causal effect of IR on t-PA antigen could be influenced by PAI-1 expression. Since we did not measure PAI-1 concentration, t-PA activity, or differentiate between total and inhibitor-bound t-PA, we were not able to characterize the mechanisms in detail. Despite its use as a fibrinolytic drug, raised t-PA levels are a marker of future cardiovascular risk (33). The causal effect of IR on t-PA antigen suggested by our findings may contribute to the excess CVD risk in individuals with diabetes and requires validation in future studies.

Our study confirmed previous reports (32) of raised t-PA concentrations to be associated with elevated T2D risk and extends these reports to an association with worsening 5-year fasting glycemia. Taken together, the current study confirms t-PA’s role in IR and T2D and points to possible a possible causal pathway from IR to t-PA concentrations.
A Novel Association Between Cathepsin D and Insulin Resistance

For six of the identified protein markers, we confirm previously reported associations with IR in humans (29, 34-38). However, the seventh protein, cathepsin D, has to our knowledge not previously been linked to IR.

The lysosomal endopeptidase cathepsin D is expressed ubiquitously and its main effects include intracellular protein turnover and extracellular matrix breakdown. Altered expression of the protein has been implicated in, for example, Alzheimer’s disease, atherosclerosis, and breast cancer (39).

Raised free fatty acid levels and advanced glycation end products found in pre-diabetic states have recently been shown to enhance cathepsin D release (40-43). This may contribute to IR through mitochondrial dysfunction (42), impaired detoxification of advanced glycation end products (43), and the induction of pro-apoptotic proteins (44).

As weight gain was shown to stimulate its activity leading to adipocyte apoptosis, cathepsin D was suggested as a potential mediator between obesity and chronic adipose tissue inflammation (44), an important contributor to IR (2). The observed strong association between cathepsin D and IR in the present study may be the result of the deranged intracellular homeostasis resulting from lipotoxicity and inflammation in insulin resistant states. Our MR study did not support a causal effect in any of the two directions and we cannot exclude that the association we identified could be due to unmeasured confounding.
Cathepsin D as a possible mediator between overweight, inflammation, and metabolic disease may be amenable to drug targeting and recent advances have been made in the field of cancer (45). Although we found a strong link between cathepsin D and prevalent IR, we failed to detect an association with 10-year diabetes risk, nor with 5-year worsening hyperglycemia. Future studies on the implication of cathepsin D in diabetes are needed.

**Protein Biomarkers for Future Diabetes and Hyperglycemia**

We confirmed the previously reported association of IL-1ra concentrations with T2D risk. IL-1ra competitively inhibits interleukin 1 from binding to its receptor, thereby suppressing its pro-inflammatory effects. In a retrospective analysis unadjusted for baseline glucose levels, IL-1ra was elevated up to 13 years prior to T2D diagnosis (46), which is in agreement with the association with 10-year incidence of T2D. Adjustment for fasting glucose resulted in a positive but insignificant association in the present study. Carstensen *et al.* (46) argued against adjusting for fasting glucose on the basis of it forming part of incident T2D. Although initially, recombinant IL-1ra agonists improved glycemic and inflammatory measures in T2D patients (47), long-term benefits have yet to be demonstrated and concerns about cardiovascular side effects from increased IL-1ra levels have been raised (48).

**Clinical Implications**

The addition of IR biomarkers to the FORS score did not improve prediction. However, the number of events in our study was moderate and larger studies are needed for more
precise estimates. The observed associations with IR support the prospective validation of
the assay for translating targeted proteomics into the diabetes care practice, but no direct
clinical implications in the short term should arise from our findings. However, the
identification of cathepsin D as an IR risk protein in this proteomics study in large
community samples suggests potential benefits of applying this technology to biomarker
discovery in the clinical setting and for other pathologies.

Limitations

Both cohorts are demographically homogeneous and consist of elderly persons, thus
limiting generalizability to other ethnic and age groups. A fasting blood sample-based
proxy measure (HOMA-IR) was used in this study, which does not provide a perfect
reflection of the physiology of IR. The scale of the proteomics assay is not readily
convertible to absolute concentrations for comparisons with previous studies. Whilst we
attempted to reduce bias in MR modeling by, e.g., sensitivity analysis for pleiotropy of
the genetic IR score, our findings are limited by the lack of statistical power for the IR-
biomarker section. Since for cathepsin D and HGF, the genetic variant used as IV maps
to the biomarker’s coding region, possible false signals due to interference with assay
antibody binding that could have resulted in an invalid instrument cannot be excluded.
Finally, the assay used focused on proteins associated with CVD and/or inflammation
and was not specifically targeted towards metabolism. An assay targeted directly towards
diabetes candidate proteins may have revealed additional findings.
Conclusion

We found evidence of a causal effect of insulin resistance on t-PA antigen concentrations, which could be part of the explanation of the excess risk of CVD in the diabetic population. We further identified cathepsin D as a novel potential biomarker for IR and demonstrated the application of high-efficiency targeted proteomics for diabetes risk assessment.
Acknowledgements

The authors thank Dr Liisa Byberg, Uppsala University, for assistance with calculation of the comorbidity score in ULSAM (Liisa Byberg reports no relevant conflicts of interest and did not receive additional funding for her assistance).

Author Contributions

TF and EI conceived the study. CN implemented, and CN, TF, JS, SG, EI planned the statistical analysis. CN wrote the draft manuscript; TF, EI, JS, SG, and LL revised it. EI, LL, VG, and JS participated in data acquisition. TF is the guarantor of the study.

Funding

This study was supported by the Swedish Research Council (VR; grant no. 2012-1397), the Swedish Heart-Lung Foundation (20140422), the Swedish Diabetes Foundation (Diabetesfonden; grant no. 2013-024), the Knut och Alice Wallenberg Foundation, the European Research Council (ERC Starting Grant), and Uppsala University. Genotyping was performed by the SNP&SEQ Technology Platform in Uppsala.

Competing Interests

TF has received honorarium for lecturing from MSD (Merck). All other authors report no competing interests.
REFERENCES


32. Eliasson MC, Jansson JH, Lindahl B, Stegmayr B. High levels of tissue plasminogen activator (tPA) antigen precede the development of type 2 diabetes in a longitudinal


ONLINE SUPPLEMENTAL MATERIALS

**Online Supplemental Figure 1. Directed acyclic graph to inform confounder choice.**
A hypothetical model of causal pathways in the association between protein biomarkers and IR was constructed with the dagitty2.2 software (http://dagitty.net/dags.html#) based on a literature review.

**Online Supplemental Figure 2. Forest plot to explore potential pleiotropy of the genetic IR score.** To explore possible pleiotropy of the genetic IR score used to test the causal effect of IR on t-PA, we extracted IR-increasing allele counts for each SNP in ULSAM and PIVUS. Following cohort-, age-, and sex-adjusted meta-analysis, heterogeneity in MR analysis findings based on the individual ten SNPs was assessed by plotting the respective IV estimators and 95% CI. The Forest plot indicates a lack of significant heterogeneity, which makes a violation of the pleiotropy assumption less likely as an explanation of our positive results.

**Online Supplemental Tables.** Tables 1-3 detail the results of the IV analysis used to infer causality. Supplemental Table 1 gives details on the genetic instrument used for HOMA-IR, and Supplemental Tables 2 and 3 list IV analysis results for causal effects of IR on biomarkers and biomarkers on IR, respectively.
### Table 1. Baseline cohort characteristics of non-diabetic participants in PIVUS-70 and ULSAM-77

<table>
<thead>
<tr>
<th>Variable</th>
<th>PIVUS</th>
<th>N</th>
<th>ULSAM</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of women</td>
<td>423 (51.1%)</td>
<td>827</td>
<td>0</td>
<td>540</td>
</tr>
<tr>
<td>Age [years]</td>
<td>70.2 ± 0.2</td>
<td>827</td>
<td>77.6 ± 0.8</td>
<td>540</td>
</tr>
<tr>
<td>BMI [kg/m²]</td>
<td>26.7 ± 4.1</td>
<td>827</td>
<td>26.0 ± 3.4</td>
<td>540</td>
</tr>
<tr>
<td>C-reactive protein [mg/l]</td>
<td>0.62 ± 0.9</td>
<td>827</td>
<td>0.67 ± 1.0</td>
<td>540</td>
</tr>
<tr>
<td>(ln-transformed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waist circumference [cm]</td>
<td>90.0 ± 11.0</td>
<td>827</td>
<td>94.6 ± 9.6</td>
<td>540</td>
</tr>
<tr>
<td>Fasting glucose [mmol/l]</td>
<td>5.5 ± 0.6</td>
<td>827</td>
<td>5.5 ± 0.6</td>
<td>540</td>
</tr>
<tr>
<td>Fasting insulin [mU/l]</td>
<td>8.3 ± 5.0</td>
<td>827</td>
<td>9.1 ± 8.0</td>
<td>540</td>
</tr>
<tr>
<td>Variable</td>
<td>Mean</td>
<td>SD</td>
<td>No.</td>
<td>SD</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>---------</td>
<td>--------</td>
<td>-----</td>
<td>--------</td>
</tr>
<tr>
<td>HOMA-IR</td>
<td>2.1 ± 1.3</td>
<td>827</td>
<td>2.3 ± 2.1</td>
<td>540</td>
</tr>
<tr>
<td>Systolic blood pressure</td>
<td>148.9 ± 22.5</td>
<td>823</td>
<td>150.3 ± 20.0</td>
<td>532</td>
</tr>
<tr>
<td>Diastolic blood pressure</td>
<td>78.4 ± 10.0</td>
<td>823</td>
<td>81.1 ± 9.8</td>
<td>532</td>
</tr>
<tr>
<td>Triglycerides [mmol/l] (ln-transformed)</td>
<td>0.1 ± 0.4</td>
<td>824</td>
<td>0.2 ± 0.4</td>
<td>539</td>
</tr>
<tr>
<td>No. of subjects with comorbidities</td>
<td>160 (19.3%)</td>
<td>827</td>
<td>208 (38.5%)</td>
<td>540</td>
</tr>
<tr>
<td>No. of subjects with a 1st-degree relative with T2D</td>
<td>115 (13.9%)</td>
<td>827</td>
<td>100 (18.5%)</td>
<td>540</td>
</tr>
<tr>
<td>Length of follow-up [years]</td>
<td>10.05 ± 0.17</td>
<td>827</td>
<td>9.13 ± 0.63</td>
<td>540</td>
</tr>
</tbody>
</table>
Expressed as mean ± standard deviation, or number of cases for categorical variables

BMI – body mass index; T2D – type 2 diabetes mellitus; HOMA-IR – homeostasis model assessment – insulin resistance

*Assessed in either arm at rest using the routine sphygmomanometer technique
**Table 2.** Linear regression analysis results for biomarker associations with lnHOMA-IR, adjusted for age, sex, body mass index, waist circumference, ln-C-reactive protein, comorbidity, and storage time.

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>PIVUS (n = 827)</th>
<th>ULSAM (n = 540)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β (95% CI)</td>
<td>P value</td>
</tr>
<tr>
<td>Leptin</td>
<td>0.27 (0.22, 0.33)</td>
<td>1.66x10^{-21}</td>
</tr>
<tr>
<td>t-PA</td>
<td>0.11 (0.07, 0.14)</td>
<td>5.97x10^{-9}</td>
</tr>
<tr>
<td>Renin</td>
<td>0.12 (0.08, 0.15)</td>
<td>4.22x10^{-11}</td>
</tr>
<tr>
<td>IL-1ra</td>
<td>0.12 (0.08, 0.16)</td>
<td>1.09x10^{-9}</td>
</tr>
<tr>
<td>HGF</td>
<td>0.15 (0.12, 0.19)</td>
<td>2.28x10^{-17}</td>
</tr>
<tr>
<td>Cathepsin D</td>
<td>0.15 (0.11, 0.18)</td>
<td>1.41x10^{-16}</td>
</tr>
<tr>
<td>FABP-4</td>
<td>0.16 (0.08, 0.17)</td>
<td>2.20x10^{-8}</td>
</tr>
</tbody>
</table>

Beta coefficients and 95% CI express the change in lnHOMA-IR associated with an SD-unit increase in normalized protein expression value. Raw P values are given for each association and all proteins shown in this table are significant at the 5% false discovery rate. Proteins with between 1-15% values below LOD in PIVUS were Protein S100-A12.
(13%), CD40 ligand (12%), TNF-related apoptosis-inducing ligand (9%), P-selectin glycoprotein ligand 1 (5%), caspase 8 (4%), leptin (4%), TNF-related activation-induced cytokine (3%), matrix metalloproteinase 3 (2%), pappalysin-1 (2%), fatty acid binding protein 4 (1%), and TNF ligand superfamily member 14 (1%); in ULSAM, the only protein was leptin (5%).

CI – confidence interval; FAPB-4 – fatty acid binding protein 4; HGF – hepatocyte growth factor; IL-1ra – interleukin-1 receptor antagonist; t-PA – tissue plasminogen activator
Table 3. Cox regression results for 10-year incident T2D and logistic regression results for 5-year risk of worse glycemia

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>10-year T2D risk (111 incident events)</th>
<th>5-year worse glycemia risk (203 incident events)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95% CI)</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td>Leptin</td>
<td>1.39 (1.00, 1.95)</td>
<td>1.02 (0.78, 1.33)</td>
</tr>
<tr>
<td>t-PA</td>
<td>1.30 (1.03, 1.65)</td>
<td>1.23 (1.02, 1.48)</td>
</tr>
<tr>
<td>Renin</td>
<td>0.86 (0.68, 1.08)</td>
<td>0.90 (0.75, 1.08)</td>
</tr>
<tr>
<td>IL-1ra</td>
<td>1.28 (1.03, 1.59)</td>
<td>1.04 (0.86, 1.25)</td>
</tr>
<tr>
<td>HGF</td>
<td>1.21 (0.98, 1.51)</td>
<td>0.98 (0.81, 1.18)</td>
</tr>
<tr>
<td>Cathepsin D</td>
<td>1.23 (0.99, 1.53)</td>
<td>0.99 (0.83, 1.19)</td>
</tr>
<tr>
<td>FABP-4</td>
<td>1.32 (0.99, 1.76)</td>
<td>0.94 (0.75, 1.17)</td>
</tr>
</tbody>
</table>

Adjusted hazard ratios (HR; adjusted for age, sex, body mass index, waist circumference, ln-C-reactive protein, storage time, and cohort) and odds ratios (OR) associated with an SD-unit increase in normalized protein expression value.
CI – confidence interval; FAPB-4 – fatty acid binding protein 4; HGF – hepatocyte growth factor; IL-1ra – interleukin-1 receptor antagonist; t-PA – tissue plasminogen activator.
FIGURE LEGENDS

Fig. 1. Flowchart illustrating the design of the study. P-values (P) were assessed at the 5% false discovery rate (FDR). Single nucleotide polymorphism (SNP) genetic variants were used as instrumental variable (IV).
Supplement

Protein biomarkers for insulin resistance and type 2 diabetes risk in two large community cohorts

Nowak et al.
Supplementary Table 1. Insulin resistance genetic risk score components and association with lnHOMA-IR

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Chromosome</th>
<th>Position (Genome Reference Consortium Human Build 38)</th>
<th>IR increasing allele</th>
<th>other allele</th>
<th>SNP score - lnHOMA-IR association in MAGIC*</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4846565</td>
<td>LYPLAL1</td>
<td>1</td>
<td>219548762</td>
<td>G</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>rs10195252</td>
<td>GRB14</td>
<td>2</td>
<td>164656581</td>
<td>T</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>rs2943645</td>
<td>IRS1</td>
<td>2</td>
<td>226234464</td>
<td>T</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>rs17036328</td>
<td>PPARG</td>
<td>3</td>
<td>12348985</td>
<td>T</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>rs3822072</td>
<td>FAM13A1</td>
<td>4</td>
<td>88820118</td>
<td>A</td>
<td>G</td>
<td>$\beta = 0.012 \pm 0.001, F=71.619, P = 2.6E-17$</td>
</tr>
<tr>
<td>rs6822892</td>
<td>PDGFC</td>
<td>4</td>
<td>156813523</td>
<td>A</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>rs4865796</td>
<td>ARL15</td>
<td>5</td>
<td>53976834</td>
<td>A</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>rs459193</td>
<td>ANKRD55/MAP3K1</td>
<td>5</td>
<td>56510924</td>
<td>G</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>rs2745353</td>
<td>RSPO3</td>
<td>6</td>
<td>127131790</td>
<td>T</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>rs731839</td>
<td>PEPD</td>
<td>19</td>
<td>33408159</td>
<td>G</td>
<td>A</td>
<td></td>
</tr>
</tbody>
</table>

F - F-statistic associated with the regression coefficient, IR - insulin resistance, P - P-value, SNP - single nucleotide polymorphism

$\beta \pm SE$ indicates the per-allele effect on lnHOMA-IR  *SNP effects based on data from PMID 20081858
Supplementary Table 2. Instrumental variable analysis for insulin resistance causally affecting biomarkers

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>Association of IR genetic risk score with biomarker in PIVUS/ULSAM*</th>
<th>IV estimator***</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β ± SE**</td>
<td>P-value</td>
</tr>
<tr>
<td>Leptin</td>
<td>-0.017 ± 0.012</td>
<td>0.174</td>
</tr>
<tr>
<td>t-PA</td>
<td>0.037 ± 0.014</td>
<td>0.008</td>
</tr>
<tr>
<td>Renin</td>
<td>0.005 ± 0.014</td>
<td>0.687</td>
</tr>
<tr>
<td>IL-1ra</td>
<td>-0.025 ± 0.014</td>
<td>0.079</td>
</tr>
<tr>
<td>HGF</td>
<td>0.010 ± 0.014</td>
<td>0.482</td>
</tr>
<tr>
<td>Cathepsin D</td>
<td>0.011 ± 0.014</td>
<td>0.410</td>
</tr>
<tr>
<td>FABP-4</td>
<td>-0.015 ± 0.013</td>
<td>0.243</td>
</tr>
</tbody>
</table>

FABP-4 - fatty acid binding protein 4, HGF - hepatocyte growth factor, IL-1ra - interleukin-1 receptor antagonist, t-PA - tissue plasminogen activator*all 10 IR-risk SNPs were directly genotyped with the Illumina Cardio-MetaboChip array in PIVUS and ULSAM **β ± SE indicates the per-allele effect on normalized protein expression value of biomarker (SD-unit) ***β ± SE indicates the causal effect of one unit change in lnHOMA-IR on normalized protein expression value of biomarker (SD-unit)
**Supplementary Table 3. Instrumental variable analysis for biomarkers causally affecting insulin resistance**

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>SNP</th>
<th>chr</th>
<th>pos</th>
<th>Closest gene</th>
<th>EA</th>
<th>Other</th>
<th>SNP - Biomarker association</th>
<th>SNP - lnHOMA-IR association</th>
<th>IV estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>β ± SE*</td>
<td>P-value</td>
<td>r²</td>
</tr>
<tr>
<td>IL-1ra</td>
<td>rs4251961</td>
<td>2</td>
<td>113116890</td>
<td>IL-1RN</td>
<td>T</td>
<td>C</td>
<td>0.082 ± 0.009</td>
<td>2.8E-21</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>rs6759676</td>
<td>2</td>
<td>113078771</td>
<td>IL1F10</td>
<td>C</td>
<td>T</td>
<td>0.075 ± 0.009</td>
<td>1.7E-17</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Combined score</td>
<td>0.079 ± 0.006</td>
<td>5.9E-35</td>
</tr>
<tr>
<td>Published GWAS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>β ± SE*</td>
<td>P-value</td>
<td>r²</td>
</tr>
<tr>
<td>HGF</td>
<td>rs5745687</td>
<td>7</td>
<td>81729735</td>
<td>HGF</td>
<td>C</td>
<td>T</td>
<td>0.099 ± 0.011</td>
<td>3.6E-19</td>
<td>2.1%</td>
</tr>
<tr>
<td>t-PA</td>
<td>rs9399599</td>
<td>6</td>
<td>147382163</td>
<td>STXB5</td>
<td>T</td>
<td>A</td>
<td>0.032 ± 0.004</td>
<td>2.9E-14</td>
<td>0.3%</td>
</tr>
<tr>
<td></td>
<td>rs3136739</td>
<td>8</td>
<td>42347562</td>
<td>POLB</td>
<td>A</td>
<td>G</td>
<td>0.063 ± 0.010</td>
<td>1.3E-09</td>
<td>0.2%</td>
</tr>
<tr>
<td></td>
<td>rs7301826</td>
<td>12</td>
<td>13086556</td>
<td>STX2</td>
<td>C</td>
<td>T</td>
<td>0.027 ± 0.004</td>
<td>1.0E-09</td>
<td>0.3%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Combined score</td>
<td>0.032 ± 0.003</td>
<td>7.0E-32</td>
</tr>
<tr>
<td>PIVUS/AM GWAS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>β ± SE*</td>
<td>P-value</td>
<td>r²</td>
</tr>
<tr>
<td>Cathepsin D</td>
<td>rs17571</td>
<td>11</td>
<td>1761364</td>
<td>CTSD</td>
<td>G</td>
<td>A</td>
<td>0.708 ± 0.059</td>
<td>9.49E-33</td>
<td>7.0%</td>
</tr>
</tbody>
</table>

GWAS - genome wide association study, HGF - hepatocyte growth factor, IL-1ra - interleukin-1 receptor antagonist, NR - not reported, t-PA - tissue plasminogen activator, chr - chromosome, pos - position (Genome Reference Consortium Human Build 38), EA - effect-allele (biomarker-increasing allele), r² – variance explained. ¶References: IL-1ra - PMID 24969107, HGF - PMID 25552591, t-PA - PMID 24578379. ¶¶proxy-SNP (r² > 0.8) for rs58561089 (position 1757375, chromosome 11, effect/non-effect alleles A/g). *β ± SE indicates the per-allele effect on protein concentration (ln-transformed for HGF and t-PA, log₂-scaled normalized protein expression value for cathepsin D) **β ± SE indicates the per-allele effect on ln-transformed HOMA-IR***β ± SE indicates the causal effect of one unit change in biomarker (ln-transformed for HGF and t-PA, SD-unit for cathepsin D) on lnHOMA-IR.
Online Supplemental Figure 1. Hypothetical causal diagram of the relationship of protein biomarkers and insulin resistance.