	Peters, Annette; Helmholtz Center Munich German Research Center for Environmental Health, Epidemiology Institute Waldenberger, Melanie; Helmholtz Center Munich German Research Center for Environmental Health, Molecular Epidemiology

Deciphering the Plasma Proteome of Type 2 Diabetes

Mohamed A. Elhadad[1,2,3] MSc., Christian Jonasson[4,5] PhD, Cornelia Huth[2,6] PhD, Rory Wilson[1,2] MSc, Christian Gieger[1,2,6] PhD, Pamela Matias[1,2,3] MSc, Harald Grallert[1,2,6] PhD, Johannes Graumann[7,8] PhD, Valerie Gailus-Durner[9] PhD, Wolfgang Rathmann[6,10] MD, Christine von Toerne[11] PhD, Stefanie M. Hauck[11] PhD, Wolfgang Koenig[3,12,13] MD, FRCP, FESC, FACC, FAHA, Moritz F. Sinner[3,14] MD, MPH, Tudor I Oprea[15,16,17] MD, PhD, Karsten Suhre[18] PhD, Barbara Thorand[2,6] PhD, Kristian Hveem[4,5] PhD, Annette Peters[2,3,6,19] PhD, Melanie Waldenberger[1,2,3] PhD

1. Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.
2. Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
3. German Research Center for Cardiovascular Disease (DZHK), Partner site Munich Heart Alliance, Germany
4. K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, NTNU - Norwegian University of Science and Technology, Trondheim, Norway
5. HUNT Research Center, Department of Public Health, NTNU - Norwegian University of Science and Technology, Levanger, Norway
6. German Center for Diabetes Research (DZD), München-Neuherberg, Ingolstädter Landstr. 1, 85764, Neuherberg, Germany
7. Biomolecular Mass Spectrometry, Max Planck Institute for Heart and Lung Research, Ludwigstrasse 43, Bad Nauheim 61231, Germany
8. The German Centre for Cardiovascular Research (DZHK), Partner Site Rhine-Main, Max Planck Institute for Heart and Lung Research, Bad Nauheim, Germany
9. German Mouse Clinic, Institute of Experimental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
10. Institute of Biometrics and Epidemiology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany
11. Research Unit Protein Science, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

12. Deutsches Herzzentrum München, Technische Universitat München, München, Germany
13. Institute of Epidemiology and Medical Biometry, University of Ulm, Ulm Germany
14. Department of Medicine I, University Hospital Munich, Ludwig-Maximilians-University, Munich, Germany
15. Department of Internal Medicine and UNM Comprehensive Cancer Center, University of New Mexico School of Medicine, Albuquerque, NM, USA
16. Department of Rheumatology and Inflammation Research, Institute of Medicine, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden
17. Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
18. Weill Cornell Medicine-Qatar, Education City, PO Box 24144, Doha, Qatar
19. Institute of Medical Information Sciences, Biometry and Epidemiology, Ludwig-Maximilians-University, Munich, Germany

Corresponding Authors:

Dr. Melanie Waldenberger

Ingolstaedter Landstr. 1, 85764, Neuherberg, Germany.

waldenberger@helmholtz-muenchen.de

Phone: +49 89 3187 1270, Fax: +49 89 3187 4567

Mohamed A. Elhadad

mohamed.elhadad@helmholtz-muenchen.de

Phone: +49 89 3187 3557, Fax: +49 89 3187 4567

Word count: 5026, three tables, three figures and supplementary materials.

Abstract:

With an estimated prevalence of 463 million affected, type 2 diabetes represents a major challenge to health care systems worldwide. Analyzing the plasma proteomes of individuals with type 2 diabetes may illuminate hitherto unknown functional mechanisms underlying disease pathology. We assessed the associations between type 2 diabetes and >1000 plasma proteins in the KORA (Cooperative health research in the Region of Augsburg) F4 cohort (n=993, 110 cases), with subsequent replication in the HUNT3 (Third wave of the Nord-Trøndelag Health Study) cohort (n=940, 149 cases). We computed logistic regression models adjusted for age, sex, BMI, smoking status and hypertension. Additionally, we investigated associations with incident type 2 diabetes and performed two-sample bi-directional Mendelian randomization (MR) analysis to prioritize our results. Association analysis of prevalent type 2 diabetes revealed 24 replicated proteins, of which eight are novel. Proteins showing association with incident type 2 diabetes were aminoacylase-1, growth hormone receptor, and insulin-like growth factor binding protein-2. Aminoacylase-1 was associated with both prevalent and incident type 2 diabetes. MR analysis yielded nominally significant causal effects of type 2 diabetes on cathepsin Z and rennin, both known to have roles in the pathophysiological pathways of cardiovascular disease, and of sex hormone-binding globulin on type 2 diabetes. In conclusion, our high-throughput proteomics study replicated previously reported type 2 diabetes-protein associations, and identified new candidate proteins possibly involved in the pathogenesis of type 2 diabetes.

Introduction

Type 2 diabetes is a significant cause of morbidity and mortality, with an estimated worldwide prevalence of 463 million patients, half of whom are undiagnosed (1). It is a complex multifactorial disease characterized by an interplay of both genetic and non-genetic factors, leading to insulin resistance and hyperinsulinemia (1; 2). Moreover, type 2 diabetes causes widespread microvascular and macrovascular complications resulting in significant health care expenditure (1).

The proteomics of type 2 diabetes, the investigation of a set of proteins within different tissues of diabetic animal models and the comparison of diabetic patients to healthy controls, has enabled the discovery of new protein-type 2 diabetes associations (3-5). Examples of associations include adiponectin (3), leptin (5) and insulin like growth factor binding protein 2 (IGFBP-2) (4). Of particular clinical interest is the study of type 2 diabetes associations with plasma proteins, which reflect systemic effects and may serve as predictive biomarkers (3; 5-7).

The integration of genetic and proteomic knowledge has provided new insight into the pathophysiology of type 2 diabetes. The best example is Mendelian randomization (MR), a method used to infer causality in observational study settings (4; 8). Previous MR studies of biomarkers and type 2 diabetes have suggested causal protective roles for proteins like adiponectin, beta-carotene, N-terminal pro B-type natriuretic peptide and sex hormone binding globulin (SHBG), as well as causal harmful roles of delta-6 desaturase and ferritin (7).

Here we use a highly multiplexed aptamer-based proteomics platform to analyze the associations between prevalent type 2 diabetes and 1095 plasma proteins in the KORA study. We replicate our results in the independent HUNT study. We investigate associations with incident type 2 diabetes using follow-up data from KORA and HUNT.

Moreover, we test the performance of our newly discovered biomarkers to predict incident type 2 diabetes when added to an adapted version of the updated German diabetes risk score (GDRS) (9). We then evaluate these newly identified proteins using the protein-protein interaction resource STRING (10). Finally, we applied two-sample bi-directional MR analysis (11) to assess causality and prioritize the newly discovered relationships.

Study design and methods

Study populations

KORA cohort

The KORA study (Cooperative Health Research in the Region of Augsburg) comprises independent samples from Augsburg, Southern Germany (12). In the current study, we used a subsample of 1000 individuals randomly selected from the participants of the KORA-F4 survey (N = 3080; performed 2006–2008) with deep phenotyping data (N = 1800) (13). Detailed clinical and socio-demographic information was collected. Data from the KORA-FF4 survey (performed 2013–2014) represents the 7-year follow-up of KORA-F4. The ethics committee of the Bavarian Medical Association reviewed and approved the study and all participants gave written informed consent.

HUNT cohort

The Nord-Trøndelag Health Study (HUNT) is a prospective population-based cohort from Nord-Trøndelag County in Norway (14). We used the HUNT3 survey (performed 2006–2008, N=1117 with proteomics measurements) for the validation of the KORA study results. The HUNT study collected detailed socio-demographic and clinical information. We used linked primary care and hospital registries for information on diabetes status at 9 years follow-up. All study participants provided written informed consent.

Proteomics measurement

Proteins were measured in fasting and non-fasting plasma samples in KORA and HUNT respectively using the SOMAscan platform as described previously (13; 15). In summary, plasma and bead-coupled aptamers, each of which has a high affinity

toward a specific protein, are incubated. After washing steps, bead-bound proteins are biotinylated and complexes comprising biotinylated target proteins and fluorescence-labelled aptamers are photocleaved off the bead support and pooled. Following recapture on streptavidin beads and further washing steps, aptamers are eluted and quantified as a proxy to protein concentration by hybridization to custom arrays of aptamer-complementary oligonucleotides. Based on standard samples included on each plate, the resulting raw intensities are processed using a data analysis work flow including hybridization normalization, median signal normalization and signal calibration to control for inter-plate differences (16). Raw intensities are reported in relative florescence units.

In KORA, one sample failed SOMAscan quality control, leaving 999 samples for analysis. Of the 1,129 SOMAmer probes (SOMAscan assay V3.2) twenty-nine failed SOMAscan quality control. We additionally removed the five probes recommended by the SOMAscan assay change log issued on December 22, 2016, leaving 1,095 probes for analysis. For replication, we used the HUNT probes that passed quality control.

Definition of outcome and model covariates

In KORA, type 2 diabetes was defined as self-reported disease validated by the responsible physician or medical chart review, or as current use of glucose-lowering medication. All participants without known diabetes were assigned to receive a standard 75 g oral glucose tolerance test (OGTT) (3). Prevalent type 2 diabetes refers to participants with the disease at the time of blood sample collection and incident refers to those developing type 2 diabetes after that time point within a 7- and 9-year follow-up period in KORA and HUNT respectively.

In HUNT, prevalent type 2 diabetes was self-reported, which we validated using clinical data from hospitals and primary care registries using the International

Classification of Diseases 10th Revision (ICD-10) code (E11) and the International Classification of Primary Care 2nd Edition (ICPC-2) code (T90). We identified incident cases of type 2 diabetes from the same registries using identical codes.

We classified participants of both cohorts who participated in leisure time physical activity for at least one hour per week as physically active (more details are available in the supplementary materials). Current hypertension was defined in KORA as having a systolic blood pressure ≥140 mmHg, diastolic blood pressure ≥90 mmHg, and/or use of antihypertensive medication. In HUNT, we used hospital and primary care data and ICD-10 codes (I10-I15) and ICPC-2 codes (K86 or K87) to identify participants with hypertension.

Drugs were assessed in KORA by asking the participants to bring the packages of their medication and supplements during their visit to the study center. Using a database software (17), medications were identified using ATC-codes, medication identifier bar code or product name.

Statistical analysis

Preprocessing of the quality controlled SOMAscan data was the same in both cohorts and involved log2 transformation and (0,1)-standardization by subtracting the per cohort mean and dividing by the per cohort standard deviation (SD) to allow easier interpretation of the odds ratios per SD of the protein.

Proteome wide analysis

Using logistic regression, we ran two proteome wide analyses in KORA: associations between proteins with prevalent type 2 diabetes, and with incident type 2 diabetes. For each of the two outcomes we ran one model per protein, i.e. 1095 models per outcome, and adjusted for the potential confounders age, sex, body mass index (BMI),

smoking status and current hypertension at baseline. We then replicated the results in HUNT using the same model. We excluded participants from both cohorts with missing values for the confounders, which led to the sample sizes of 993 and 940 for KORA and HUNT, respectively. For the analysis with incident type 2 diabetes we further excluded all participants with prevalent type 2 diabetes, resulting in sample sizes of 881 and 794 for KORA and HUNT, respectively. We used the false discovery rate (FDR) Benjamini–Hochberg method separately for the outcomes to account for multiple testing. An association was considered statistically significant at FDR<0.05.

We replicated significant results in HUNT using the same model. We considered proteins replicated at FDR<0.05, with FDR calculated based on the number of significant proteins in KORA.

To examine whether anti-diabetic drug intake influenced the replicated associations, we ran sensitivity analyses by including the drugs of interest as confounders one at a time.

Data analytics of replicated proteins

The candidate proteins were processed via the Pharos (18) platform, experimental Gene Ontology (GO) (19) terms, KEGG (20) pathways, human disease association data from the GWAS Catalog (21), OMIM (22) and the text-mining DISEASES platform (23), as well as phenotype data from the corresponding mouse orthologue knockouts (24). We mined these resources for data on the potential associations between the candidate proteins and type 2 diabetes.

Prediction of incident diabetes

We applied a biomarker discovery strategy to investigate whether proteins significantly associated with incident type 2 diabetes in KORA could be used for prediction of

incident type 2 diabetes. These were ten proteins, of which one failed QC in HUNT. We used KORA as a training dataset and HUNT as a test dataset, and used an adapted version of GDRS, a diabetes risk score that was trained in 21845 participants of the prospective EPIC-Potsdam study with a mean follow-up time of 7 years, as a benchmark (9). More details on the GDRS score are available in the supplementary materials. Adaptation of the GDRS risk score was necessary as some of the variables were missing from one or both cohorts, and was performed as follows: We defined smoking status using only information on current and former smoking per se without regard to the amount of cigarettes smoked. We used the average of the original GDRS score weights for each smoking category to represent our combined categories (former: (15+45)/2 = 30, current: (23+77)/2 = 50). Family history of diabetes was defined in KORA as having at least one parent or sibling with diabetes and in HUNT as having at least one parent, sibling or a child with diabetes. We calculated the risk of a positive family history by averaging the original GDRS risk scores of having one parent, both parents or at least one sibling with type 2 diabetes ((56+106+48)/3 = 70). Our final adapted GDRS score was calculated as follows: 5.1×age in years + 7.6×waist circumference in cm – 2.7×height in cm + 47×hypertension status - 2×physical activity (at least one hour per week) + 30×former smoking + 50×current smoking + 70×family history of type 2 diabetes.

We tested prediction performance of all models using the area under the receiver operating characteristic (ROC) curve (AUC), and applied the DeLong test to compare AUCs of nested models.

First, we added the proteins to the adapted GDRS model and applied the least absolute shrinkage and selection operator (LASSO) (25) for model selection in KORA. LASSO shrinks the sum of the absolute values of the regression coefficients, forcing

some to be set to zero, thus performing a form of model selection. The LASSO lambda was chosen by cross-validation using the squared-error for Gaussian models. The GDRS was calculated in each cohort and used as a score which was fixed by setting its penalty factor to zero to prevent any shrinkage by LASSO. We then compared the performance of LASSO protein-extended-GDRS model to the adapted GDRS model. We assessed the calibration of the LASSO-selected model using calibration plots (26). Moreover, we tested the performance of the proteins as single predictors on top of the adapted GDRS model.

Mendelian randomization

We attempted to infer causality of the replicated proteins associated with type 2 diabetes by applying two-sample bi-directional Mendelian randomization. Figure 1 shows the summary of the pipeline for the causal inference analysis. In summary, we extracted single nucleotide polymorphisms (SNPs) as instrumental variables (IV) from published genome wide association study (GWAS) summary statistics of European ancestry if they passed the Bonferroni threshold of $p<5e-8$. We extracted the IVs from the meta-analysis of type 2 diabetes GWAS studies by Xue et al. (N = 455607) and the GWAS studies of SOMAscan measured proteins; Sun et al. (N = 3301), Suhre et al. (N = 1000) and Emilsson et al. (N = 5457) for proteins (13; 27-29). We identified ambiguous palindromic SNPs, defined as SNPs with A/T or G/C alleles and an effect allele frequency around 0.5 using the cutoff points defined by the “TwoSampleMR” R-package (30). We replaced these with a proxy SNP, defined as a SNP with r2 exceeding 0.85 with the SNP in question, when available, or excluded them from further analyses (31). We then clumped the SNPs, which implies removing SNPs in linkage disequilibrium with the lead SNP using the r2 cut-off 0.001. We did not manually prune the final list of IVs. Further, IVs selected for proteins needed to be in

cis, i.e. within one megabase of the protein-coding gene as per the Human Genome Assembly GRCh37.p13.

We proceeded to extract the results of these IVs or of one of their proxies from the outcome's GWAS. For proteins, priority was given to results from Sun et al. (28) because of the larger sample size, followed by Suhre et al. (13), dependent on availability.

We used the Wald ratio to check for causality (32). In cases of more than one IV, we used the random effects model of the inverse variance weighted meta-analysis to combine the Wald ratio estimates of all IVs (8; 32). For sensitivity analyses whenever there was more than one IV, we ran the MR-Egger regression model to look for horizontal pleiotropy in our causal models (33), leave-one-out analysis and forest plots to identify outliers among these IVs that would be driving the results in a certain direction, and examined scatter plots to check for outliers.

Analytical steps are summarized in the flowchart Figure S1 presented in supplementary materials. All analyses were done in R version 3.5.1 (The R Foundation for Statistical Computing). For MR analysis, the "TwoSampleMR" R-package version 0.4.22 was employed (30).

Data and Resource Availability

Informed consents given by KORA study participants do not cover data posting in public databases. However, the KORA data is available given approval of online requests at the KORA Project Application Self-Service Tool (https://epi.helmholtz-muenchen.de/). The HUNT data can be accessed given approval of applications to HUNT Research Centre (http://www.ntnu.edu/hunt/data). The data used in the Mendelian randomization analysis are publicly available and can be accessed through:

http://cnsgenomics.com/data.html (Xue et al. at), http://www.phpc.cam.ac.uk/ceu/proteins/ (Sun et al. at), http://proteomics.gwas.eu (Suhre et al.) and www.sciencemag.org/content/361/6404/769/suppl/DC1 (Emilsson et al.). Example code for the analytic steps of the manuscript could be accessed at https://github.com/maelhadad/T2D_SOMAscan_Proteomics.

Results

Descriptive statistics of the study populations

Table 1 and supplementary table S1 show the baseline characteristics of both cohorts and their follow-up subsets respectively. HUNT participants were on average older, and comprised more men.

Association results of plasma proteins with type 2 diabetes

The proteome wide analysis with prevalent type 2 diabetes yielded 85 FDR significant proteins (Supplementary Table S2), of which 24 successfully replicated in HUNT (Table 2 & Figure 2A). Of these, osteomodulin was most strongly associated (based on KORA p-value) with an odds ratio (OR) per SD increase in protein level of 0.61 (95% CI 0.47 - 0.77) in KORA and of 0.65 (0.53 - 0.79) in HUNT. Among the positively associated proteins, peptide YY had the strongest association (OR = 1.34; 95% CI = (1.1 - 1.62) in KORA and OR = 1.58; 95% CI = (1.32 - 1.92) in HUNT).

To assess whether the proteome panel was associated with future type 2 diabetes, we performed a proteome wide analysis with incident type 2 diabetes using the same model, which yielded ten FDR significant protein associations (Supplementary Table S3). Of these, aminoacylase-1, growth hormone receptor and IGFBP-2 replicated in HUNT (Table 3 & Figure 2B). Adiponectin failed quality control in HUNT and thus replication was not possible. Among the replicated proteins, aminoacylase-1 showed the strongest association (OR = 1.78; 95% CI = (1.34 - 2.37) in KORA and OR = 1.6; 95% CI = (1.26 - 2.05) in HUNT). Interestingly, aminoacylase-1 overlapped between the replicated results of both prevalent and incident type 2 diabetes.

Additionally, we assessed the concordance of the effect estimates across the cohorts. Of 85 KORA FDR-significant proteins associated with prevalent type 2 diabetes, only

seven had different effect directions, but none of these was nominally significant in HUNT (Figure 3A). For incident type 2 diabetes, two proteins showed opposite effect directions, neither of these reaching nominal statistical significance (Figure 3B).

Overlap with known type 2 diabetes genetic and protein associations

To assess the overlap between our results and known type 2 diabetes associations, we compared our results to gene-based results described by Xue et al (27). Furthermore, we compared our replicated proteins to protein lists of interest published by the Human Diabetes Proteome Project, namely the 1000 diabetes-related proteins, the human islet of Langerhans proteome, the rodent beta-cell proteome and the human blood glycated proteome (34). Of the 26 unique replicated proteins, 18 overlapped with at least one list. Eight proteins have not been previously found to be related to type 2 diabetes (Supplementary Table S4).

Data analytics of replicated proteins

Supplementary table S5 shows information extracted from Pharos for our replicated proteins. Alpha-L-iduronidase, cathepsin A and cathepsin Z shared the same lysosomal pathway association according to KEGG (20).

Investigating potential effects of drugs on type 2 diabetes-protein associations in KORA

None of the replicated protein-incident type 2 diabetes associations showed loss of significance when adjusting for any of the investigated drugs. On the other hand, three of the replicated associations with prevalent type 2 diabetes lost statistical significance when adjusting for anti-diabetic medication intake (Supplementary Table S6 and Figure S2). All of the associations retained the same direction of effect apart from PYY,

which showed an opposite effect after adjusting for anti-diabetic medication, and, more specifically, metformin.

Prediction of incident type 2 diabetes

Starting with the nine proteins associated with incident type 2 diabetes in KORA available in HUNT, we evaluated whether a subset of them selected using LASSO would improve the predictive performance of the adapted GDRS benchmark model (9). LASSO selected five proteins, namely transforming growth factor beta-receptor type 3 (TGFbR3), tartrate-resistant acid phosphatase type 5, pappalysin-1, afamin and scavenger receptor cysteine-rich type 1 protein M130 (sCD163). The LASSO selected protein-enhanced model showed improvement in both KORA and HUNT (GDRS-protein extended AUC = 0.84 with 95% CI = (0.79 – 0.89) and 0.67 with 95% CI = (0.61 – 0.72); GDRS-only AUC = 0.77 with 95% CI = (0.71 – 0.83) and 0.66 with 95% CI = (0.60 – 0.72), respectively); however, according to the Delong test the AUC improvement in HUNT was not statistically significant (p-value = 0.72, Supplementary Figure S3). The calibration plot of the LASSO-selected model in HUNT yielded an intercept of 0.23 and a slope of 0.53 (Supplementary Figure S3). The intercept of the calibration plot examines the difference of means of predicted and observed risk. In HUNT, it is higher than zero thus showing higher observed type 2 diabetes cases in HUNT as those predicted. This could be attributed to longer follow up in HUNT; 9 years compared to 7 years in KORA and to the fact that HUNT is older than KORA and therefore have more cases of type 2 diabetes. The slope of calibration is 0.53 in HUNT, which indicates a possible overfitting of the model or the need for coefficient shrinkage in HUNT that could also be attributed to the heterogeneity between the study populations in terms of patient characteristics and outcome definition. The training dataset used a gold standard screening to define type 2 diabetes where HUNT did not

apply similar definition and would therefore have hidden cases and measurement error. Therefore, the outcome being predicted for HUNT (and defined by KORA) is slightly different to the outcome observed.

We further tested the performance of individual proteins as predictors of incident type 2 diabetes in KORA and validated our models in HUNT (Supplementary Figure S4). The following proteins showed relatively similar performance in both cohorts: aminoacylase-1 (KORA-AUC= 0.78 with 95% CI = (0.73 – 0.84); HUNT-AUC= 0.71 with 95% CI = (0.65 – 0.77)), growth hormone receptor (KORA-AUC= 0.77 with 95% CI = (0.71 – 0.83); HUNT-AUC= 0.70 with 95% CI = (0.64 – 0.76)) and IGFBP-2 (KORA-AUC= 0.78 with 95% CI = (0.72 – 0.84); HUNT-AUC= 0.73 with 95% CI = (0.68 – 0.79)).

Mendelian randomization analysis of replicated plasma proteins and type 2 diabetes in KORA

Using up to 120 SNPs as genetic instruments, we investigated if type 2 diabetes had a causal effect on the 26 replicated proteins from both the prevalence and incidence analyses (Supplementary Table S7 and Figure S5). For cathepsin Z (MR-IVW-Beta= 0.13; p-value = 2.00e-03) and renin (MR-IVW-Beta= 0.08; p-value = 3.15e-02), a nominally significant causal effect of prevalent type 2 diabetes was observed, each with the same direction of effect as their observational results. MR-Egger analyses to test for the presence of horizontal pleiotropy showed no significant results for either protein (intercept p-values of 0.17 and 0.1 for cathespin Z and renin, respectively). Tests and plots to check for outliers in the instrumental variables showed no significant aberrations (Supplementary Figure S6 and S7).

We also ran MR to investigate if any of the proteins had a causal effect on type 2 diabetes. We analyzed 13 proteins for which we found independent cis acting IVs

(Supplementary Table S8 and Figure S5). We observed a nominally significant causal effect of SHBG on type 2 diabetes, with the same direction of effect as its observed association (MR-Wald-Beta = -0.09; p-value = 2.95e-02). None of the associations for either direction survived Bonferroni multiple testing correction.

Discussion

We report a proteome wide analysis of type 2 diabetes in KORA and replication in HUNT using aptamer-based affinity proteomics. Our analysis yielded 26 unique replicated significant protein associations. Of these, 24 replicated exclusively with prevalent type 2 diabetes, two replicated exclusively with incident type 2 diabetes, and aminoacylase-1 replicated with both.

Aminoacylase-1 is a zinc-dependent peptidase involved in amino acid metabolism (35). The protein has not been described in the context of type 2 diabetes before, but has been reported to be overexpressed in obese liver tissue, thus linking it to obesity and inflammation (36). A further study found aminoacylase-1 to be down-regulated in obese omental fat, which the authors hypothesized to be due to adipocyte dysfunction caused by obesity (35). Moreover, aminoacylase-1 is associated with arginine production according to KEGG (20). Plasma levels of arginine were found to be higher in type 2 diabetes patients (37).

In addition to aminoacylase-1, incident type 2 diabetes results included an inverse association with IGFBP-2 and a positive association with growth hormone receptor. IGFBP-2 was reported to have type 2 diabetes-protective effects and has been shown to reverse hyperglycemia in insulin and leptin deficiency (38). These associations highlight the role of the growth hormone axis in the early pathophysiology of type 2 diabetes. Both growth hormone and insulin like growth factor-1 are known to play roles in the insulin receptor cascade, leading to insulin resistance (39).

The analysis of prevalent type 2 diabetes confirmed previously known proteomic associations like gelsolin (40), renin (41), SHBG (42) and hepatocyte growth factor receptor, and revealed promising new candidate proteins, including osteomodulin,

matrilin-2, Wnt inhibitory factor-1 (WIF1), tumor necrosis factor inducible gene 6 protein (TNFAIP6), cerebral dopamine neurotrophic factor (CDNF), RGM domain family member B, TGFbR3 and SLIT & NTRK-like protein 5, which were downregulated in type 2 diabetes cases, and lysosomal protective protein, galectin-3 binding protein (LGALS3BP) and peptide YY (PYY), which were upregulated.

Our results overlap and complement results of mass-spectrometry studies on obesity. Plasma levels of apolipoprotein B, LGALS3BP and SHBG were found to be altered by sustained weight loss (43) and gastric by-pass surgery induced weight loss (44) with the latter affecting also plasma protease C1 inhibitor, complement C2 and gelsolin.

New protein associations with prevalent type 2 diabetes included proteins previously reported in association with complications of type 2 diabetes. Increased circulating levels of LGALS3BP were linked to non-alcoholic fatty liver disease (45) and acute venous thrombosis (46) and TGFbR3 was reported to be associated with diabetic nephropathy (47). TNFAIP6 and CDNF were shown to have protective effects, while WIF1, TGFBR3 and PYY were reported to have harmful effects, in the development and progress of cardiovascular atherosclerotic diseases (48-52). Along this line, members of the complement family like plasma protease C1 inhibitor and complement C2 were downregulated and upregulated in our results respectively, and proteins from the renin-angiotensin and kallikrein-kinin systems included the upregulated renin and downregulated kallikrein-7.

Although our study cohorts were different regarding the fasting status of their samples, most proteins (78 out of 85 for prevalent and 8 out of 10 for incident type 2 diabetes) showed concordant effects between cohorts, while none of the non-concordant proteins were statistically significant in the replication (Figure 3). Nonetheless, fasting has significant metabolic consequences that are expected to be reflected in the

plasma proteome and could have contributed to non-replication in HUNT. However, there are multiple other potential explanations for the non-replication, perhaps differing from one protein to another. Importantly, while plasma protein levels differ between fasting and non-fasting samples, this does not necessarily match the variance in the protein levels caused by the disease status. As such, disease-related variance would still be apparent despite differences in fasting status. For example, some of our examined proteins were reported to show differences in their levels according to fasting status, like SHBG (53), PYY (53) and sCD163 (54), yet their associations with disease status were replicated in our study. Of the proteins that failed replication in HUNT, MMP2 (53) and Pappalysin-1 (55) have been found to be affected by food intake. However, their effect sizes were similar in both cohorts suggesting fasting status may not be the primary reason for non-replication for most proteins.

Sensitivity analyses into the potential effect of drugs on the type 2 diabetes plasma protein associations showed loss of significance of some associations after adjusting for anti-diabetic medication. The effect direction of all the resultant associations remained the same, except for PYY, which showed a change of direction after adjusting for metformin intake; however, as its effect estimate was not significant after adjustment it is difficult to draw any conclusions from this.

Additionally, we evaluated the significant proteins' ability to predict incident type 2 diabetes. The protein-extended models showed improved performance over the adapted GDRS benchmark model (9) in both the KORA discovery and the HUNT replication, although the improvement was very small and not statistically significant (p-value = 0.72) for the latter. Moreover, we tested the performances of individual proteins on top of the adapted benchmark model. The best performances in the replication cohort came from aminoacylase-1, growth hormone receptor and IGFBP2,

each of which achieved approximately equal performance in HUNT compared to KORA, results which warrant validation in clinical trials using commercially available ELISA kits. As the KORA samples were taken from individuals in a fasting state (>= 8 hours) and HUNT samples were taken non-fasting, these results seem to indicate that fasting status is largely irrelevant with regard to type 2 diabetes prediction for these candidate biomarkers. However, fasting may potentially be relevant for other markers, since the AUC was much smaller in HUNT compared to KORA for some of the other measured biomarkers in combination with the GDRS score.

Our investigations into the causal framework governing the relation between plasma proteins and type 2 diabetes showed suggestive harmful causal effects of SHBG on type 2 diabetes. SHBG has been previously reported to be associated with type 2 diabetes (42) and may be implicated in the development of insulin resistance (42). We demonstrated it to be negatively associated with type 2 diabetes, a causal direction suggested by the MR analysis as well.

Causal inference analysis showed suggestive causal effects of type 2 diabetes on both cathepsin Z and renin. In line with previous observations, we demonstrated renin to be positively associated with type 2 diabetes in both observational and MR analysis results (41). The association is an indicator of the upregulated renin-angiotensin-aldosterone system, which is activated in obesity and type 2 diabetes, thus contributing to CVD complications (41; 56). Cathepsin Z is a member of the peptidase C1 family that plays a role in lysosomal function, which might explain its connection to diabetes through beta-cell failure driven by lysosomal degradation (57).

Study strengths

We applied a high throughput proteomics platform on samples from population-based cohorts for our analyses, which enabled us to test a large number of proteins with a

wide concentration range and to generalize our results to our samples' respective populations. We used samples from plasma, which is easily accessible, and is the usual medium of biomarkers. Additionally, the plasma proteome reflects on the levels of proteins originating from a broad range of tissues, thus giving us insight into systemic pathways. Finally, we were able to test for the causal relationship in both directions using publically available data on genetic associations with both type 2 diabetes and proteins.

Study limitations

We are aware of several limitations to our study. First, aptamer-based proteomics is susceptible to potential probe cross-reactivity and non-specific binding (28; 29). However, we verified that none of the proteins identified have been flagged for such issues (validation data presented in supplementary materials methods, table S10, and figures S10 and S11). Due to the lack of OGTT data in HUNT, the rigorous definition of type 2 diabetes employed in KORA could not be extended and the discrepancy in fasting status between the cohorts may have contributed to the limited replication of our results. Our prediction models do not reflect the dynamic changes in the proteome, which would require a more detailed investigation. This is also true for the Mendelian randomization results, which reflect the lifelong genetic risk rather than point change in single protein levels in relation to disease status. Although, there is an overlap between the participants of the genetic datasets used for type 2 diabetes and proteins through KORA, none of the associations -tested using such data- were significant.

Conclusion

Our proteome wide analysis of type 2 diabetes replicated known associations and revealed novel candidate proteins. Associations with incident type 2 diabetes included aminacylase-1, which overlapped with prevalent type 2 diabetes associations. New associations with prevalent type 2 diabetes included TNFAIP6, CDNF, WIF1, TGFBR3 and PYY, all of which are thought to play a role in the development of cardiovascular complications like atherosclerosis. Mendelian randomization suggested a causal role of SHBG on type 2 diabetes, which is in line with previous observational and MR analysis results. It also suggested a causal effect of type 2 diabetes on both cathepsin Z and renin, both of which are known to play a role in type 2 diabetes complications. Our results offer insight into proteins involved in the pathogenesis of type 2 diabetes and its complications, proteins that could be valuable drug targets for all levels of prevention.

Acknowledgments

Funding. 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. This work was supported by a grant (WA 4081/1-1) from the German Research Foundation and by the German Federal Ministry of Education and Research (BMBF) within the framework of the EU Joint Programming Initiative ‘A Healthy Diet for a Healthy Life’ (DIMENSION grant number 01EA1902A). This work was further supported by the Biomedical Research Program at Weill Cornell Medicine in Qatar, a program funded by the Qatar Foundation. The HUNT3 study was funded by the Norwegian Ministry of Health, Norwegian University of Science and Technology and

Norwegian Research Council, Central Norway Regional Health Authority, the Nord-Trondelag County Council and the Norwegian Institute of Public Health. M.A.E. is supported by the German Center for Cardiovascular Research (DZHK). K.S. is supported by Qatar National Research Fund (QNRF) grant no. NPRPC11-0115-180010. T.I.O. is supported by NIH grants U24 CA224370, U24 TR002278 and U01 CA239108.

Conflict of Interest. CJ has received personal fees for research consultancy work from Pfizer and Bayer outside of the submitted work; T.I.O. has received honoraria or consulted for Abbott, AstraZeneca, Chiron, Genentech, Infinity Pharmaceuticals, Merz Pharmaceuticals, Merck Darmstadt, Mitsubishi Tanabe, Novartis, Ono Pharmaceuticals, Pfizer, Roche, Sanofi and Wyeth outside of the submitted work. All other authors report no potential conflicts of interest relevant to this article.

Author Contributions. M.A.E and M.W. designed the study. M.A.E. analyzed the data, interpreted the results, wrote and revised the manuscript. R.W., V.G.D. and T.I.O. helped with the analyses. M.W., R.W., C.J., K.H., C.H., C.G., P.M., H.G., J.G., C.vT., S.M.H., W.R., W.K., M.F.S., K.S, B.T. and A.P. were involved in the data collection, data management and preparation of their corresponding cohort. All authors contributed to the writing of the manuscript, critically reviewed it and approved the final version for submission. M.A.E and M.W. are the guarantors of this work and, as such, had full access to all the data in the study and take full responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. International Diabetes Federation. IDF Diabetes Atlas, 9th edn. Brussels, Belgium: 2019. Available at: http://www.diabetesatlas.org. 2019;
2. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nature reviews Endocrinology 2018;14:88-98
3. Huth C, von Toerne C, Schederecker F, de las Heras Gala T, Herder C, Kronenberg F, Meisinger C, Rathmann W, Koenig W, Waldenberger M, Roden M, Peters A, Hauck SM, Thorand B. Protein markers and risk of type 2 diabetes and prediabetes: a targeted proteomics approach in the KORA F4/FF4 study. European Journal of Epidemiology 2019;34:409-422
4. Beijer K, Nowak C, Sundström J, Ärnlöv J, Fall T, Lind L. In search of causal pathways in diabetes: a study using proteomics and genotyping data from a cross-sectional study. Diabetologia 2019;
5. Nowak C, Sundstrom J, Gustafsson S, Giedraitis V, Lind L, Ingelsson E, Fall T. Protein Biomarkers for Insulin Resistance and Type 2 Diabetes Risk in Two Large Community Cohorts. Diabetes 2016;65:276-284
6. von Toerne C, Huth C, de Las Heras Gala T, Kronenberg F, Herder C, Koenig W, Meisinger C, Rathmann W, Waldenberger M, Roden M, Peters A, Thorand B, Hauck SM. MASP1, THBS1, GPLD1 and ApoA-IV are novel biomarkers associated with prediabetes: the KORA F4 study. Diabetologia 2016;59:1882-1892
7. Abbasi A. Mendelian randomization studies of biomarkers and type 2 diabetes. Endocrine connections 2015;4:249-260
8. Zheng J, Baird D, Borges MC, Bowden J, Hemani G, Haycock P, Evans DM, Smith GD. Recent Developments in Mendelian Randomization Studies. Current epidemiology reports 2017;4:330-345
9. Paprott R, Mühlenbruch K, Mensink GBM, Thiele S, Schulze MB, Scheidt-Nave C, Heidemann C. Validation of the German Diabetes Risk Score among the general adult population: findings from the German Health Interview and Examination Surveys. BMJ Open Diabetes Research & Care 2016;4:e000280
10. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic acids research 2019;47:D607-d613
11. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 2013;37:658-665
12. Holle R, Happich M, Lowel H, Wichmann HE, Group MKS. KORA--a research platform for population based health research. Gesundheitswesen 2005;67 Suppl 1:S19-25
13. Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, Sarwath H, Thareja G, Wahl A, DeLisle RK, Gold L, Pezer M, Lauc G, El-Din Selim MA, Mook-Kanamori DO, Al-Dous EK, Mohamoud YA, Malek J, Strauch K, Grallert H, Peters A, Kastenmüller G, Gieger C, Graumann J. Connecting genetic risk to disease end points through the human blood plasma proteome. Nature Communications 2017;8:14357
14. Krokstad S, Langhammer A, Hveem K, Holmen TL, Midthjell K, Stene TR, Bratberg G, Heggland J, Holmen J. Cohort Profile: the HUNT Study, Norway. International journal of epidemiology 2013;42:968-977
15. Ganz P, Heidecker B, Hveem K, Jonasson C, Kato S, Segal MR, Sterling DG, Williams SA. Development and Validation of a Protein-Based Risk Score for Cardiovascular Outcomes Among Patients With Stable Coronary Heart Disease. Jama 2016;315:2532-2541
16. Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN, Carter J, Dalby AB, Eaton BE, Fitzwater T, Flather D, Forbes A, Foreman T, Fowler C, Gawande B, Goss M, Gunn M, Gupta S, Halladay D, Heil J, Heilig J, Hicke B, Husar G, Janjic N, Jarvis T, Jennings S, Katilius E, Keeney TR, Kim N, Koch TH, Kraemer S, Kroiss L, Le N, Levine D, Lindsey W, Lollo B, Mayfield W, Mehan M, Mehler R, Nelson SK, Nelson M, Nieuwlandt D, Nikrad M, Ochsner U, Ostroff RM, Otis M, Parker T, Pietrasiewicz S,

Resnicow DI, Rohloff J, Sanders G, Sattin S, Schneider D, Singer B, Stanton M, Sterkel A, Stewart A, Stratford S, Vaught JD, Vrkljan M, Walker JJ, Watrobka M, Waugh S, Weiss A, Wilcox SK, Wolfson A, Wolk SK, Zhang C, Zichi D. Aptamer-based multiplexed proteomic technology for biomarker discovery. PloS one 2010;5:e15004
17. Mühlberger N, Behrend C, Stark RG, Holle R. *Datenbankgestützte Online-Erfassung von Arzneimitteln im Rahmen gesundheitswissenschaftlicher Studien - Erfahrungen mit der IDOM-Software*. München, Urban & Fischer, 2003
18. Nguyen DT, Mathias S, Bologa C, Brunak S, Fernandez N, Gaulton A, Hersey A, Holmes J, Jensen LJ, Karlsson A, Liu G, Ma'ayan A, Mandava G, Mani S, Mehta S, Overington J, Patel J, Rouillard AD, Schurer S, Sheils T, Simeonov A, Sklar LA, Southall N, Ursu O, Vidovic D, Waller A, Yang J, Jadhav A, Oprea TI, Guha R. Pharos: Collating protein information to shed light on the druggable genome. Nucleic acids research 2017;45:D995-d1002
19. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic acids research 2018;47:D330-D338
20. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. Nucleic acids research 2019;47:D590-d595
21. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, Suveges D, Vrousgou O, Whetzel PL, Amode R, Guillen JA, Riat HS, Trevanion SJ, Hall P, Junkins H, Flicek P, Burdett T, Hindorff LA, Cunningham F, Parkinson H. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic acids research 2019;47:D1005-d1012
22. Amberger JS, Bocchini CA, Scott AF, Hamosh A. OMIM.org: leveraging knowledge across phenotype-gene relationships. Nucleic acids research 2019;47:D1038-d1043
23. Pletscher-Frankild S, Pallejà A, Tsafou K, Binder JX, Jensen LJ. DISEASES: Text mining and data integration of disease–gene associations. Methods 2015;74:83-89
24. Koscielny G, Yaikhom G, Iyer V, Meehan TF, Morgan H, Atienza-Herrero J, Blake A, Chen C-K, Easty R, Di Fenza A, Fiegel T, Grifiths M, Horne A, Karp NA, Kurbatova N, Mason JC, Matthews P, Oakley DJ, Qazi A, Regnart J, Retha A, Santos LA, Sneddon DJ, Warren J, Westerberg H, Wilson RJ, Melvin DG, Smedley D, Brown SDM, Flicek P, Skarnes WC, Mallon A-M, Parkinson H. The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. Nucleic acids research 2013;42:D802-D809
25. Tibshirani R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society Series B (Methodological) 1996;58:267-288
26. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21:128-138
27. Xue A, Wu Y, Zhu Z, Zhang F, Kemper KE, Zheng Z, Yengo L, Lloyd-Jones LR, Sidorenko J, Wu Y, Agbessi M, Ahsan H, Alves I, Andiappan A, Awadalla P, Battle A, Beutner F, Bonder Marc J, Boomsma D, Christiansen M, Claringbould A, Deelen P, Esko T, Favé M-J, Franke L, Frayling T, Gharib S, Gibson G, Hemani G, Jansen R, Kähönen M, Kalnapenkis A, Kasela S, Kettunen J, Kim Y, Kirsten H, Kovacs P, Krohn K, Kronberg-Guzman J, Kukushkina V, Kutalik Z, Lee B, Lehtimäki T, Loeffler M, Marigorta UM, Metspalu A, Milani L, Müller-Nurasyid M, Nauck M, Nivard M, Penninx B, Perola M, Pervjakova N, Pierce B, Powell J, Prokisch H, Psaty B, Raitakari O, Ring S, Ripatti S, Rotzschke O, Ruëger S, Saha A, Scholz M, Schramm K, Seppälä I, Stumvoll M, Sullivan P, Teumer A, Thiery J, Tong L, Tönjes A, van Dongen J, van Meurs J, Verlouw J, Völker U, Võsa U, Yaghootkar H, Zeng B, McRae AF, Visscher PM, Zeng J, Yang J, e QC. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. Nature Communications 2018;9:2941
28. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, Burgess S, Jiang T, Paige E, Surendran P, Oliver-Williams C, Kamat MA, Prins BP, Wilcox SK, Zimmerman ES, Chi A, Bansal N, Spain SL, Wood AM, Morrell NW, Bradley JR, Janjic N, Roberts DJ, Ouwehand WH, Todd JA, Soranzo N, Suhre K, Paul DS, Fox CS, Plenge RM, Danesh J, Runz H, Butterworth AS. Genomic atlas of the human plasma proteome. Nature 2018;558:73-79

29. Emilsson V, Ilkov M, Lamb JR, Finkel N, Gudmundsson EF, Pitts R, Hoover H, Gudmundsdottir V, Horman SR, Aspelund T, Shu L, Trifonov V, Sigurdsson S, Manolescu A, Zhu J, Olafsson O, Jakobsdottir J, Lesley SA, To J, Zhang J, Harris TB, Launer LJ, Zhang B, Eiriksdottir G, Yang X, Orth AP, Jennings LL, Gudnason V. Co-regulatory networks of human serum proteins link genetics to disease. Science (New York, NY) 2018;361:769-773
30. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith G, Gaunt TR, Haycock PC. The MR-Base platform supports systematic causal inference across the human phenome. eLife 2018;7
31. Hartwig FP, Davies NM, Hemani G, Davey Smith G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. International journal of epidemiology 2017;45:1717-1726
32. Teumer A. Common Methods for Performing Mendelian Randomization. Front Cardiovasc Med 2018;5:51-51
33. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International journal of epidemiology 2015;44:512-525
34. Topf F, Schvartz D, Gaudet P, Priego-Capote F, Zufferey A, Turck N, Binz PA, Fontana P, Wiederkehr A, Finamore F, Xenarios I, Goodlett D, Kussmann M, Bergsten P, Sanchez JC. The Human Diabetes Proteome Project (HDPP): From network biology to targets for therapies and prevention. Translational Proteomics 2013;1:3-11
35. Perez-Perez R, Garcia-Santos E, Ortega-Delgado FJ, Lopez JA, Camafeita E, Ricart W, Fernandez-Real JM, Peral B. Attenuated metabolism is a hallmark of obesity as revealed by comparative proteomic analysis of human omental adipose tissue. Journal of proteomics 2012;75:783-795
36. Caira S, Iannelli A, Sciarrillo R, Picariello G, Renzone G, Scaloni A, Addeo P. Differential representation of liver proteins in obese human subjects suggests novel biomarkers and promising targets for drug development in obesity. Journal of enzyme inhibition and medicinal chemistry 2017;32:672-682
37. Cao Y-F, Li J, Zhang Z, Liu J, Sun X-Y, Feng X-F, Luo H-H, Yang W, Li S-N, Yang X, Fang Z-Z. Plasma Levels of Amino Acids Related to Urea Cycle and Risk of Type 2 Diabetes Mellitus in Chinese Adults. Frontiers in Endocrinology 2019;10
38. Hedbacker K, Birsoy K, Wysocki RW, Asilmaz E, Ahima RS, Farooqi IS, Friedman JM. Antidiabetic effects of IGFBP2, a leptin-regulated gene. Cell metabolism 2010;11:11-22
39. Kim S-H, Park M-J. Effects of growth hormone on glucose metabolism and insulin resistance in human. Ann Pediatr Endocrinol Metab 2017;22:145-152
40. Khatri N, Sagar A, Peddada N, Choudhary V, Chopra BS, Garg V, Garg R, Ashish. Plasma gelsolin levels decrease in diabetic state and increase upon treatment with F-actin depolymerizing versions of gelsolin. Journal of diabetes research 2014;2014:152075
41. Joseph Joshua J, Echouffo Tcheugui Justin B, Effoe Valery S, Hsueh Willa A, Allison Matthew A, Golden Sherita H. Renin-Angiotensin-Aldosterone System, Glucose Metabolism and Incident Type 2 Diabetes Mellitus: MESA. Journal of the American Heart Association 2018;7:e009890
42. Le TN, Nestler JE, Strauss JF, 3rd, Wickham EP, 3rd. Sex hormone-binding globulin and type 2 diabetes mellitus. Trends Endocrinol Metab 2012;23:32-40
43. Geyer PE, Wewer Albrechtsen NJ, Tyanova S, Grassl N, Iepsen EW, Lundgren J, Madsbad S, Holst JJ, Torekov SS, Mann M. Proteomics reveals the effects of sustained weight loss on the human plasma proteome. Mol Syst Biol 2016;12:901
44. Wewer Albrechtsen NJ, Geyer PE, Doll S, Treit PV, Bojsen-Moller KN, Martinussen C, Jorgensen NB, Torekov SS, Meier F, Niu L, Santos A, Keilhauer EC, Holst JJ, Madsbad S, Mann M. Plasma Proteome Profiling Reveals Dynamics of Inflammatory and Lipid Homeostasis Markers after Roux-En-Y Gastric Bypass Surgery. Cell Syst 2018;7:601-612.e603
45. Niu L, Geyer PE, Wewer Albrechtsen NJ, Gluud LL, Santos A, Doll S, Treit PV, Holst JJ, Knop FK, Vilsboll T, Junker A, Sachs S, Stemmer K, Muller TD, Tschop MH, Hofmann SM, Mann M. Plasma

proteome profiling discovers novel proteins associated with non-alcoholic fatty liver disease. Mol Syst Biol 2019;15:e8793
46. DeRoo EP, Wrobleski SK, Shea EM, Al-Khalil RK, Hawley AE, Henke PK, Myers DD, Jr., Wakefield TW, Diaz JA. The role of galectin-3 and galectin-3-binding protein in venous thrombosis. Blood 2015;125:1813-1821
47. Reeves WB, Andreoli TE. Transforming growth factor β contributes to progressive diabetic nephropathy. Proceedings of the National Academy of Sciences 2000;97:7667-7669
48. Watanabe R, Sato Y, Ozawa N, Takahashi Y, Koba S, Watanabe T. Emerging Roles of Tumor Necrosis Factor-Stimulated Gene-6 in the Pathophysiology and Treatment of Atherosclerosis. Int J Mol Sci 2018;19:465
49. Zhang G-L, Wang L-H, Liu X-Y, Zhang Y-X, Hu M-Y, Liu L, Fang Y-Y, Mu Y, Zhao Y, Huang S-H, Liu T, Wang X-J. Cerebral Dopamine Neurotrophic Factor (CDNF) Has Neuroprotective Effects against Cerebral Ischemia That May Occur through the Endoplasmic Reticulum Stress Pathway. Int J Mol Sci 2018;19:1905
50. Ress C, Paulweber M, Goebel G, Willeit K, Rufinatscha K, Strobl A, Salzmann K, Kedenko L, Tschoner A, Staudacher G, Iglseder B, Tilg H, Paulweber B, Kaser S. Circulating Wnt inhibitory factor 1 levels are associated with development of cardiovascular disease. Atherosclerosis 2018;273:1-7
51. Sun F, Li X, Duan W-Q, Tian W, Gao M, Yang J, Wu X-Y, Huang D, Xia W, Han Y-N, Wang J-X, Liu Y-X, Dong C-J, Zhao D, Ban T, Chu W-F. Transforming Growth Factor-β Receptor III is a Potential Regulator of Ischemia-Induced Cardiomyocyte Apoptosis. Journal of the American Heart Association 2017;6:e005357
52. Smith RM, Klein R, Kruzliak P, Zulli A. Role of Peptide YY in blood vessel function and atherosclerosis in a rabbit model. Clin Exp Pharmacol Physiol 2015;42:648-652
53. Pellis L, van Erk MJ, van Ommen B, Bakker GCM, Hendriks HFJ, Cnubben NHP, Kleemann R, van Someren EP, Bobeldijk I, Rubingh CM, Wopereis S. Plasma metabolomics and proteomics profiling after a postprandial challenge reveal subtle diet effects on human metabolic status. Metabolomics 2012;8:347-359
54. Dencker M, Gårdinger Y, Björgell O, Hlebowicz J. Effect of food intake on 92 biomarkers for cardiovascular disease. PloS one 2017;12:e0178656
55. Dencker M, Gaayrdinger Y, BjAalrgell O, Hlebowicz J. Effect of food intake on biomarkers for cardiovascular disease and inflammation analyzed with the Proseek Multiplex CVD II kit. Genetics & Molecular Research 2018;17
56. Bahtiyar G, Gutterman D, Lebovitz H. Heart Failure: a Major Cardiovascular Complication of Diabetes Mellitus. Current Diabetes Reports 2016;16:116
57. Pasquier A, Vivot K, Erbs E, Spiegelhalter C, Zhang Z, Aubert V, Liu Z, Senkara M, Maillard E, Pinget M, Kerr-Conte J, Pattou F, Marciniak G, Ganzhorn A, Ronchi P, Schieber NL, Schwab Y, Saftig P, Goginashvili A, Ricci R. Lysosomal degradation of newly formed insulin granules contributes to beta cell failure in diabetes. Nat Commun 2019;10:3312

Table 1: Baseline characteristics of the prevalent study populations.

Variable	**KORA (n=993)**	**HUNT (n=940)**	**p value***
Age † (years)	59.31 (43 - 79)	69.03 (31.6 - 99.4)	<0.001
Sex female ‡	514 (51.8 %)	245 (26.1 %)	<0.001
BMI § (kg/m2)	27.79 (4.58)	28.36 (3.96)	0.003
Waist circumference § (cm)	94.51 (13.81)	100.01 (11.01)	<0.001
Physical inactivity ‡	376 (37.9 %)	472 (49.2 %)	<0.001
Smoking ‡ Never smoker Former smoker Current smoker	 423 (42.6 %) 422 (42.5 %) 148 (14.9 %)	 234 (24.9 %) 504 (53.6 %) 202 (21.5 %)	<0.001
Family history of diabetes ‡	312 (38.1 %)	280 (31.6%)	0.005
Hypertension ‡	396 (39.9 %)	389 (41.4 %)	0.531

* Continuous variables were tested for a difference between the two populations using t-tests and categorical variables with Chi-square tests with continuity correction.

† Mean (range); ‡ number (percentage); § mean ± standard deviation.

Table 2: Results of the proteome wide analysis with prevalent type 2 diabetes in KORA, for those proteins statistically significant in the discovery (FDR<0.05) and their replicated results (FDR<0.05, same direction of effect) in HUNT sorted alphabetically. All analyses were adjusted for age, sex, body mass index (BMI), smoking status and hypertension.

Protein Full Name	Protein Short Name	UniProt ID	KORA (n = 993)			HUNT (n = 940)		
			OR (95% CI)	p value	FDR p-value	OR (95% CI)	p value	FDR p-value
Alpha-L-iduronidase	IDUA	P35475	1.48 (1.2 - 1.84)	3.04E-04	1.07E-02	1.44 (1.19 - 1.74)	1.59E-04	1.42E-03
Aminoacylase-1	Aminoacylase-1	Q03154	2.1 (1.64 - 2.71)	5.62E-09	2.05E-06	1.49 (1.22 - 1.84)	1.26E-04	1.26E-03
Apolipoprotein B	Apo B	P04114	0.48 (0.37 - 0.61)	4.19E-09	2.05E-06	0.7 (0.57 - 0.84)	2.87E-04	1.86E-03
Cathepsin Z	CATZ	Q9UBR2	1.41 (1.13 - 1.77)	2.27E-03	3.35E-02	1.33 (1.1 - 1.62)	3.20E-03	1.37E-02
Cerebral dopamine neurotrophic factor	ARMEL	Q49AH0	0.64 (0.48 - 0.82)	7.25E-04	1.85E-02	0.7 (0.55 - 0.87)	1.83E-03	8.62E-03
Complement C2	C2	P06681	2.01 (1.37 - 3.04)	6.63E-04	1.81E-02	1.47 (1.2 - 1.82)	3.03E-04	1.86E-03
Galectin-3-binding protein	LG3BP	Q08380	1.6 (1.27 - 2.01)	5.04E-05	2.51E-03	1.43 (1.2 - 1.72)	9.47E-05	1.08E-03
Gelsolin	Gelsolin	P06396	0.55 (0.43 - 0.69)	4.31E-07	9.44E-05	0.66 (0.54 - 0.81)	4.31E-05	6.88E-04
Hepatocyte growth factor receptor	Met	P08581	0.62 (0.49 - 0.78)	4.89E-05	2.51E-03	0.78 (0.65 - 0.92)	4.93E-03	1.88E-02
Kallikrein-7	Kallikrein 7	P49862	0.59 (0.46 - 0.75)	1.47E-05	1.46E-03	0.67 (0.54 - 0.82)	1.95E-04	1.56E-03
Lysosomal protective protein	Cathepsin A	P10619	1.54 (1.24 - 1.92)	8.51E-05	3.88E-03	1.32 (1.09 - 1.6)	5.48E-03	1.91E-02
Matrilin-2	MATN2	O00339	0.62 (0.49 - 0.77)	2.77E-05	2.17E-03	0.7 (0.57 - 0.86)	7.17E-04	3.82E-03
Osteomodulin	OMD	Q99983	0.61 (0.47 - 0.77)	3.89E-05	2.34E-03	0.64 (0.52 - 0.78)	1.22E-05	3.26E-04
Peptide YY	PYY	P10082	1.34 (1.1 - 1.62)	3.36E-03	4.59E-02	1.53 (1.27 - 1.86)	9.26E-06	3.26E-04
Periostin	Periostin	Q15063	0.54 (0.43 - 0.68)	1.52E-07	4.16E-05	0.75 (0.62 - 0.92)	5.50E-03	1.91E-02
Plasma protease C1 inhibitor	C1-Esterase Inhibitor	P05155	0.67 (0.53 - 0.84)	5.39E-04	1.59E-02	0.76 (0.61 - 0.93)	9.66E-03	3.22E-02
Renin	Renin	P00797	1.61 (1.32 - 1.99)	5.48E-06	6.67E-04	1.45 (1.21 - 1.74)	5.16E-05	6.88E-04

RGM domain family member B	RGMB	Q6NW40	0.64 (0.49 - 0.81)	3.52E-04	1.20E-02	0.73 (0.59 - 0.9)	3.25E-03	1.37E-02
Sex hormone-binding globulin	SHBG	P04278	0.62 (0.47 - 0.8)	2.41E-04	9.11E-03	0.63 (0.51 - 0.77)	1.12E-05	3.26E-04
SLIT and NTRK-like protein 5	SLIK5	O94991	0.6 (0.47 - 0.76)	3.81E-05	2.34E-03	0.78 (0.66 - 0.93)	4.38E-03	1.75E-02
Transforming growth factor beta receptor type 3	TGF-b R III	Q03167	0.58 (0.46 - 0.73)	4.45E-06	6.09E-04	0.74 (0.61 - 0.89)	1.40E-03	6.99E-03
Trypsin-1	Trypsin	P07477	0.63 (0.5 - 0.78)	4.06E-05	2.34E-03	0.7 (0.58 - 0.84)	2.19E-04	1.59E-03
Tumor necrosis factor-inducible gene 6 protein	TSG-6	P98066	0.58 (0.45 - 0.74)	2.27E-05	2.07E-03	0.7 (0.57 - 0.85)	3.96E-04	2.26E-03
Wnt inhibitory factor 1	WIF-1	Q9Y5W5	0.5 (0.37 - 0.66)	2.97E-06	4.75E-04	0.65 (0.54 - 0.79)	2.21E-05	4.43E-04

Table 3: Results of the proteome wide analysis with incident type 2 diabetes in KORA, for those proteins statistically significant in the discovery (FDR<0.05) and their replicated results (FDR<0.05, same direction of effect) in HUNT sorted alphabetically. All analyses were adjusted for age, sex, BMI, smoking status and hypertension.

Protein Full Name	Protein Short Name	UniProt ID	KORA (n = 881)			HUNT (n = 794)		
			OR (95% CI)	p-value	FDR p-value	OR (95% CI)	p-value	FDR p-value
Aminoacylase-1	Aminoacylase-1	Q03154	1.78 (1.34 - 2.37)	7.15E-05	1.96E-02	1.6 (1.26 - 2.04)	1.27E-04	1.14E-03
Growth hormone receptor	Growth hormone receptor	P10912	1.74 (1.31 - 2.38)	2.43E-04	3.32E-02	1.42 (1.07 - 1.88)	1.37E-02	4.11E-02
Insulin-like growth factor-binding protein 2	IGFBP-2	P18065	0.47 (0.34 - 0.65)	6.07E-06	2.22E-03	0.57 (0.42 - 0.77)	2.91E-04	1.31E-03

Figure 1: Mendelian randomization analysis flowchart; a Clumping refers to the process of selecting only the independent IVs, i.e. those that are not in linkage disequilibrium (LD) with each other, using the cutoff (LD-r2 value > 0.001); b Harmonizing the data refers to ensuring that the effects of the IV on the exposure and the outcome reflect the same strand effect.

Figure 2: Volcano plot of type 2 diabetes results in KORA, where proteins that replicated in HUNT are labeled; A) Results of the proteome wide analysis with prevalent type 2 diabetes in KORA; B) Results of the proteome wide analysis with incident type 2 diabetes in KORA.

Figure 3: Coefficient concordance between KORA and HUNT for A) prevalent type 2 diabetes and B) incident type 2 diabetes. Proteins that replicated in HUNT are labeled.

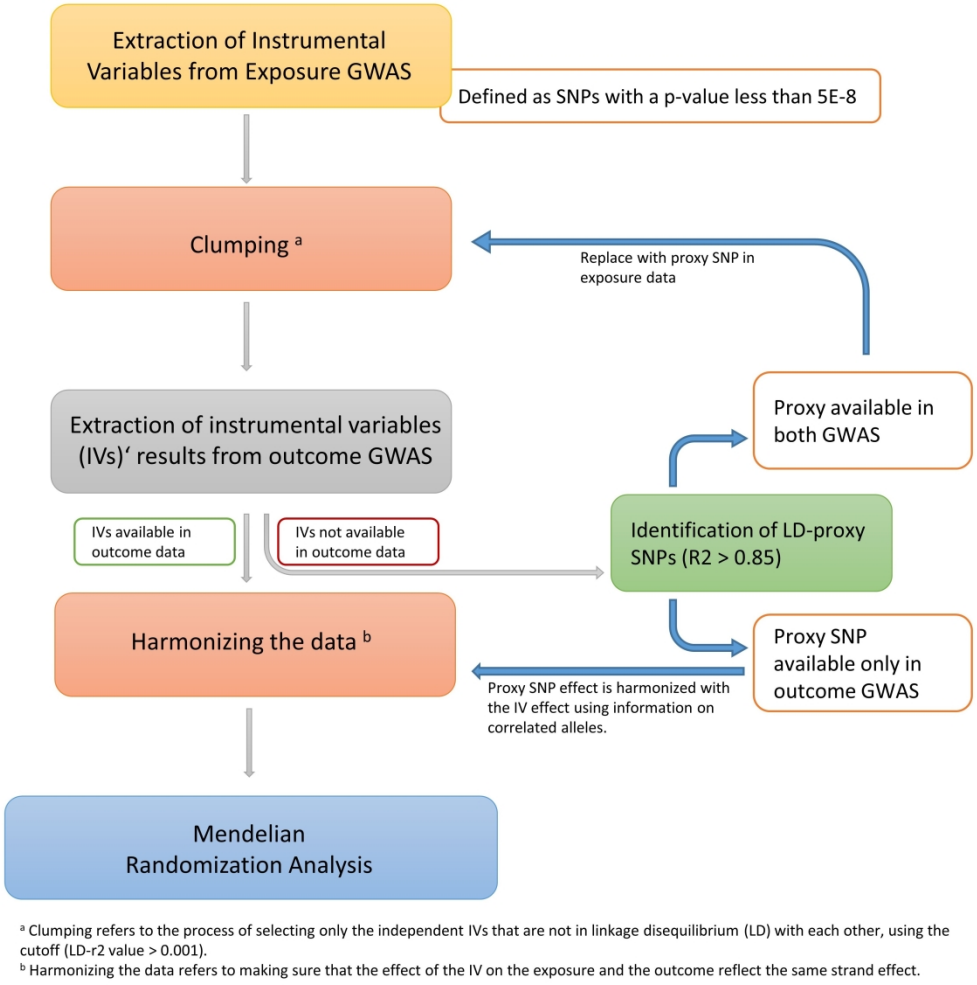

Figure 1: Mendelian randomization analysis flowchart; a Clumping refers to the process of selecting only the independent IVs, ie those that are not in linkage disequilibrium (LD) with each other, using the cutoff (LD-r2 value > 0.001); b Harmonizing the data refers to ensuring that the effects of the IV on the exposure and the outcome reflect the same strand effect.

1066x1079mm (96 x 96 DPI)

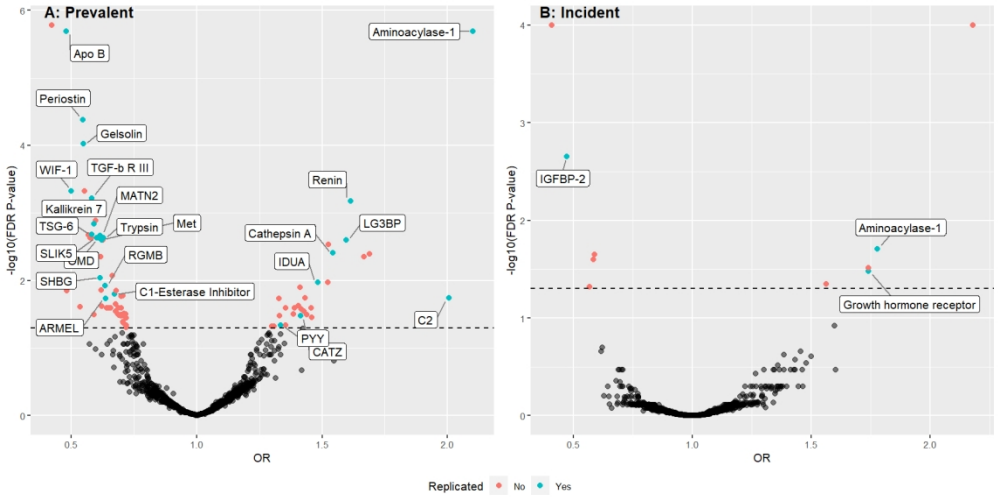

Figure 2: Volcano plot of T2D results in KORA, where replicated proteins are labeled; A) Results of the proteome wide analysis with prevalent type 2 diabetes in KORA; B) Results of the proteome wide analysis with incident type 2 diabetes in KORA.

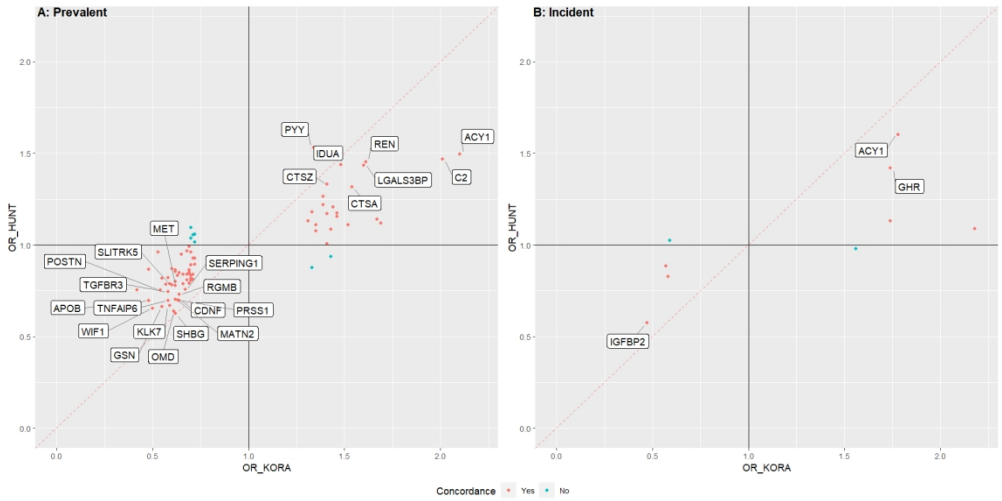

Figure 3: Coefficient concordance between KORA and HUNT for A) prevalent type 2 diabetes and B) incident type 2 diabetes. Proteins showing disagreement in effect direction between cohorts are labeled.

Supplementary material

The updated German Diabetes Risk Score

The updated German Diabetes Risk Score (GDRS) is a risk score that was trained in 21845 participants of the prospective EPIC-Potsdam study with a mean follow-up time of 7 years (1). It was validated in 3625 participants of the German National Health Interview and Examination Survey 1998 (baseline) with follow up data in the German Health Interview and Examination Survey for Adults (follow-up) (2).

The score is based on clinical phenotypes and aims to identify those at risk of developing type 2 diabetes in the German population. It is built by assigning weights to each variable, which are derived from the regression model used to develop the score (1; 2).

Physical activity definition in both cohorts

In KORA, physical activity was assessed using a questionnaire with about weekly exercise frequency and duration in summer and winter (3). Potential answers were (1) > 2 hours, (2) 1–2 hours, (3) <1 hour and (4) none. Participants with a total score < 5, obtained by summing the numbers (1)–(4) relating to winter and summer, we considered to be ‘physically active’. Those who walked at least 30 minutes a day were also considered ‘physically active’. In HUNT, we applied the same definition to maintain consistency.

Validation of the SOMAscan protein measurements

We checked the performance of the SOMAscan platform in KORA by examining the correlation between SOMAscan measured biomarkers and their overlapping counterparts measured using other techniques.

These proteins were leptin measured in plasma by ELISA from Mercodia (Stockholm, Sweden), C-reactive protein (CRP), cystatin-C measured in plasma by a high-sensitivity latex-enhanced nephelometric assay on a BN II analyzer from Dade Behring (Erlangen, Germany) and 7 proteins measured by selected reaction monitoring mass spectrometry (SRM-MS).

We used Pearson correlation to test the concordance between proteins measured with SOMAscan and other techniques in KORA (Supplementary Figure S10, S11). CRP, leptin and cystatin-C showed good correlation, with their r ranging between 0.94 for CRP and 0.75 for cystatin-C. Proteins measured by SRM-MS showed good correlation for CRP, Mannose-binding protein C (MBL2), Thrombospondin-1 (THBS1), SHBG and adiponectin (ADIPOQ) with r ranging from 0.69 to 0.85 while RBP4 and CD5L showed lower r of 0.31 and 0.55 respectively.

Furthermore, we checked the validity of the replicated proteins by examining the data from Emilsson V. et al, where the authors validated the SOMAscan measurements using mass spectrometry (4). We found information on the validation for 22 of the unique 26 replicated proteins (Supplementary Table S10).

STRING protein-protein interaction network analysis

We queried STRING (5), the protein-protein interaction server, to visually assess the relationship between these candidate proteins and other proteins connected to type 2 diabetes. We used our replicated proteins and the type 2 diabetes associated proteins curated from UniProt as seed proteins and specified that no additional proteins (interactors) should be added by STRING to the network.

Among the type 2 diabetes associated proteins in UniProt (6) (Supplementary table S5; Protein origin: UniProtKB), 19 out of 23 proteins form a complex network with 17 of our 26 replicated proteins in STRING (Supplementary Figure S12).

Table S1: Baseline characteristics of the incident cohorts with those with type 2 diabetes at baseline excluded.

Variable	**KORA (n=881)**	**HUNT (n=794)**	**p value***
Age † (years)	58.63 (43 - 75)	68.83 (31.6 - 99.4)	<0.001
Sex female ‡	464 (52.7 %)	199 (25.1 %)	<0.001
BMI § (kg/m2)	27.41 (4.41)	28.04 (3.69)	0.002
Waist circumference § (cm)	93.31 (13.55)	98.94 (10.49)	<0.001
Physical inactivity ‡	560 (63.6 %)	407 (52.2 %)	<0.001
Smoking ‡ Never smoker Former smoker Current smoker	 382 (43.4 %) 367 (41.7 %) 132 (15.0 %)	 192 (24.2 %) 430 (54.2 %) 172 (21.7 %)	<0.001
Family history of diabetes ‡	255 (34.8 %)	201 (26.8 %)	0.001
Hypertension ‡	316 (35.9 %)	304 (38.3 %)	0.331

* Continuous variables were tested for a difference between the two populations using t-tests and categorical variables with Chi-square tests with continuity correction.

† Mean (range); ‡ number (percentage); § mean ± standard deviation.

Table S2: KORA FDR significant results of prevalent type 2 diabetes in KORA, HUNT as well as their combined effect using a meta-analysis random effects model. Model was adjusted for age and sex, body mass index (BMI), smoking and hypertension.

Protein Full Name	UniProt ID	KORA (n = 993)		HUNT (n = 940)		Combined	
		OR (95% CI)	p value	OR (95% CI)	p value	OR (95% CI)	p value
72 kDa type IV collagenase	P08253	0.7 (0.56 - 0.87)	1.83E-03	0.89 (0.73 - 1.08)	2.36E-01	0.79 (0.63 - 1.01)	5.63E-02
Adiponectin	Q15848	0.57 (0.44 - 0.75)	3.96E-05				
Afamin	P43652	1.52 (1.22 - 1.92)	2.97E-04	1.11 (0.91 - 1.36)	3.20E-01	1.29 (0.95 - 1.77)	1.05E-01
Alpha-1-antichymotrypsin	P01011	0.71 (0.56 - 0.89)	3.07E-03	0.93 (0.77 - 1.11)	4.14E-01	0.82 (0.63 - 1.06)	1.35E-01
Alpha-1-antichymotrypsin complex	P01011	0.7 (0.56 - 0.88)	2.25E-03	1.09 (0.9 - 1.31)	3.49E-01	0.88 (0.57 - 1.36)	5.66E-01
Alpha-L-iduronidase	P35475	1.48 (1.2 - 1.84)	3.04E-04	1.44 (1.19 - 1.74)	1.59E-04	1.46 (1.26 - 1.68)	1.78E-07
Aminoacylase-1	Q03154	2.1 (1.64 - 2.71)	5.62E-09	1.49 (1.22 - 1.84)	1.26E-04	1.76 (1.26 - 2.46)	9.23E-04
Apolipoprotein B	P04114	0.48 (0.37 - 0.61)	4.19E-09	0.7 (0.57 - 0.84)	2.87E-04	0.58 (0.4 - 0.84)	3.63E-03
Arylsulfatase A	P15289	1.31 (1.09 - 1.57)	3.62E-03	1.13 (0.96 - 1.33)	1.37E-01	1.21 (1.05 - 1.39)	7.88E-03
Calpastatin	P20810	0.62 (0.46 - 0.82)	1.00E-03	0.86 (0.7 - 1.05)	1.44E-01	0.74 (0.54 - 1.03)	7.13E-02
Cathepsin D	P07339	1.39 (1.14 - 1.7)	1.13E-03	1.22 (1.01 - 1.48)	4.02E-02	1.3 (1.13 - 1.49)	1.90E-04
Cathepsin Z	Q9UBR2	1.41 (1.13 - 1.77)	2.27E-03	1.33 (1.1 - 1.62)	3.20E-03	1.37 (1.18 - 1.58)	2.39E-05
C-C motif chemokine 23	P55773	0.7 (0.56 - 0.88)	2.92E-03	0.81 (0.66 - 0.99)	3.99E-02	0.76 (0.65 - 0.89)	4.44E-04
CD97 antigen	P48960	0.66 (0.51 - 0.84)	1.14E-03	0.84 (0.68 - 1.02)	8.69E-02	0.75 (0.59 - 0.96)	2.03E-02
Cerebral dopamine neurotrophic factor	Q49AH0	0.64 (0.48 - 0.82)	7.25E-04	0.7 (0.55 - 0.87)	1.83E-03	0.67 (0.56 - 0.79)	4.89E-06
Chordin-like protein 1	Q9BU40	0.57 (0.43 - 0.74)	2.51E-05	0.79 (0.63 - 0.98)	2.94E-02	0.67 (0.49 - 0.93)	1.47E-02
Ciliary neurotrophic factor receptor subunit alpha	P26992	0.71 (0.57 - 0.89)	2.30E-03	0.81 (0.67 - 0.98)	3.16E-02	0.77 (0.66 - 0.88)	2.84E-04
Ck-beta-8-1	P55773	0.66 (0.53 - 0.82)	2.18E-04	0.79 (0.65 - 0.96)	1.58E-02	0.73 (0.61 - 0.86)	2.47E-04
Coagulation factor IX	P00740	1.67 (1.3 - 2.18)	1.10E-04	1.14 (0.94 - 1.41)	2.02E-01	1.37 (0.94 - 1.98)	9.88E-02
Coagulation factor IXab	P00740	1.69 (1.31 - 2.22)	9.18E-05	1.12 (0.93 - 1.37)	2.63E-01	1.36 (0.91 - 2.04)	1.34E-01
Coiled-coil domain-containing protein 80	Q76M96	0.62 (0.48 - 0.79)	1.05E-04	0.85 (0.68 - 1.07)	1.78E-01	0.73 (0.53 - 1)	5.15E-02
Collagenase 3	P45452	0.53 (0.36 - 0.76)	1.06E-03	0.96 (0.77 - 1.16)	6.89E-01	0.73 (0.41 - 1.3)	2.86E-01
Complement C2	P06681	2.01 (1.37 - 3.04)	6.63E-04	1.47 (1.2 - 1.82)	3.03E-04	1.63 (1.22 - 2.19)	9.06E-04
Complement C3b, inactivated	P01024	1.42 (1.15 - 1.77)	1.34E-03				
Complement component C9	P02748	0.72 (0.58 - 0.88)	1.71E-03	1.06 (0.87 - 1.3)	5.73E-01	0.87 (0.59 - 1.28)	4.84E-01
Cystatin-M	Q15828	0.68 (0.54 - 0.85)	8.95E-04	0.84 (0.69 - 1.02)	8.11E-02	0.76 (0.62 - 0.94)	1.01E-02
Dickkopf-like protein 1	Q9UK85	1.35 (1.14 - 1.66)	1.26E-03	1.08 (0.89 - 1.3)	4.33E-01	1.21 (0.97 - 1.51)	9.82E-02
Dickkopf-related protein 3	Q9UBP4	0.69 (0.54 - 0.87)	1.95E-03	0.86 (0.7 - 1.06)	1.70E-01	0.78 (0.62 - 0.97)	2.64E-02
Ectodysplasin-A, secreted form	Q92838	0.64 (0.49 - 0.83)	1.18E-03	0.85 (0.68 - 1.04)	1.30E-01	0.75 (0.57 - 0.98)	3.56E-02
Endoplasmic reticulum resident protein 29	P30040	1.44 (1.15 - 1.82)	1.97E-03	1.21 (1 - 1.45)	4.81E-02	1.3 (1.1 - 1.54)	2.64E-03

Endostatin	P39060	0.7 (0.57 - 0.86)	6.07E-04	1.04 (0.85 - 1.27)	7.21E-01	0.85 (0.58 - 1.26)	4.17E-01
Endothelial cell-specific molecule 1	Q9NQ30	0.62 (0.47 - 0.8)	4.30E-04	0.8 (0.62 - 1)	6.68E-02	0.71 (0.55 - 0.91)	6.96E-03
Endothelin-converting enzyme 1	P42892	0.71 (0.57 - 0.88)	1.66E-03	0.84 (0.7 - 1)	5.24E-02	0.78 (0.66 - 0.92)	3.17E-03
Ephrin type-B receptor 2	P29323	0.68 (0.55 - 0.84)	4.68E-04	0.97 (0.79 - 1.18)	7.41E-01	0.81 (0.57 - 1.15)	2.40E-01
Ephrin type-B receptor 6	O15197	0.7 (0.55 - 0.88)	2.26E-03	0.96 (0.78 - 1.16)	6.86E-01	0.82 (0.6 - 1.13)	2.25E-01
Fibroblast growth factor 19	O95750	0.6 (0.47 - 0.75)	1.19E-05	0.87 (0.72 - 1.04)	1.35E-01	0.73 (0.5 - 1.05)	8.71E-02
Fibronectin Fragment 4	P02751	0.7 (0.56 - 0.87)	1.27E-03	0.82 (0.68 - 0.99)	3.65E-02	0.76 (0.65 - 0.89)	7.03E-04
Ficolin-3	O75636	1.46 (1.15 - 1.87)	2.46E-03	1.15 (0.94 - 1.42)	1.70E-01	1.28 (1.02 - 1.62)	3.18E-02
Galectin-3-binding protein	Q08380	1.6 (1.27 - 2.01)	5.04E-05	1.43 (1.2 - 1.72)	9.47E-05	1.5 (1.3 - 1.72)	2.40E-08
Galectin-4	P56470	1.33 (1.1 - 1.59)	2.29E-03	1.18 (1.01 - 1.39)	3.36E-02	1.24 (1.1 - 1.39)	3.39E-04
GDNF family receptor alpha-2	O00451	0.63 (0.5 - 0.78)	5.02E-05	0.83 (0.69 - 1.01)	5.86E-02	0.73 (0.55 - 0.96)	2.45E-02
Gelsolin	P06396	0.55 (0.43 - 0.69)	4.31E-07	0.66 (0.54 - 0.81)	4.31E-05	0.61 (0.51 - 0.73)	1.06E-07
Hepatocyte growth factor receptor	P08581	0.62 (0.49 - 0.78)	4.89E-05	0.78 (0.65 - 0.92)	4.93E-03	0.7 (0.57 - 0.88)	1.58E-03
Iduronate 2-sulfatase	P22304	0.69 (0.54 - 0.87)	1.77E-03	0.85 (0.71 - 1.02)	7.03E-02	0.77 (0.63 - 0.95)	1.35E-02
Insulin-like growth factor-binding protein 2	P18065	0.42 (0.32 - 0.55)	1.53E-09	0.76 (0.59 - 0.97)	2.57E-02	0.57 (0.32 - 1)	5.19E-02
Interleukin-11 receptor subunit alpha	Q14626	0.58 (0.44 - 0.75)	3.23E-05	0.82 (0.67 - 1)	5.49E-02	0.7 (0.49 - 0.98)	3.98E-02
Interleukin-22 receptor subunit alpha-2	Q969J5	0.72 (0.57 - 0.9)	3.42E-03	0.93 (0.78 - 1.11)	4.07E-01	0.82 (0.64 - 1.06)	1.26E-01
Kallikrein-7	P49862	0.59 (0.46 - 0.75)	1.47E-05	0.67 (0.54 - 0.82)	1.95E-04	0.63 (0.54 - 0.74)	1.49E-08
Kynureninase	Q16719	1.35 (1.1 - 1.65)	3.44E-03	1.11 (0.92 - 1.33)	2.66E-01	1.22 (1 - 1.48)	4.53E-02
Legumain	Q99538	1.41 (1.18 - 1.74)	3.80E-04	1.17 (0.97 - 1.4)	8.96E-02	1.28 (1.07 - 1.54)	7.56E-03
Leucine carboxyl methyltransferase 1	Q9UIC8	0.48 (0.32 - 0.72)	4.68E-04	0.87 (0.69 - 1.07)	1.90E-01	0.66 (0.37 - 1.18)	1.59E-01
Lysosomal protective protein	P10619	1.54 (1.24 - 1.92)	8.51E-05	1.32 (1.09 - 1.6)	5.48E-03	1.42 (1.21 - 1.65)	1.05E-05
Matrilin-2	O00339	0.62 (0.49 - 0.77)	2.77E-05	0.7 (0.57 - 0.86)	7.17E-04	0.66 (0.57 - 0.77)	1.05E-07
Melanoma-derived growth regulatory protein	Q16674	0.69 (0.55 - 0.87)	2.09E-03	0.79 (0.65 - 0.96)	1.75E-02	0.75 (0.65 - 0.87)	1.47E-04
Muellerian-inhibiting factor	P03971	0.72 (0.57 - 0.9)	3.85E-03	1.02 (0.84 - 1.21)	8.70E-01	0.86 (0.61 - 1.21)	3.81E-01
Myoglobin	P02144	0.68 (0.53 - 0.86)	1.51E-03	0.81 (0.66 - 0.99)	4.18E-02	0.75 (0.63 - 0.89)	1.16E-03
NADPH--cytochrome P450 reductase	P16435	1.43 (1.14 - 1.78)	1.52E-03	1.08 (0.9 - 1.31)	4.00E-01	1.24 (0.94 - 1.62)	1.23E-01
Netrin receptor UNC5D	Q6UXZ4	0.55 (0.43 - 0.71)	3.04E-06	0.82 (0.69 - 0.97)	2.28E-02	0.68 (0.46 - 1)	4.98E-02
Neurexin-1-beta	P58400	0.59 (0.42 - 0.81)	1.97E-03	0.79 (0.62 - 0.98)	4.66E-02	0.7 (0.53 - 0.92)	1.21E-02
Neurogenic locus notch homolog protein 1	P46531	0.7 (0.56 - 0.88)	1.99E-03	0.83 (0.69 - 1)	5.41E-02	0.77 (0.66 - 0.91)	1.93E-03
Osteomodulin	Q99983	0.61 (0.47 - 0.77)	3.89E-05	0.64 (0.52 - 0.78)	1.22E-05	0.62 (0.54 - 0.73)	2.03E-09
Pappalysin-1	Q13219	0.71 (0.57 - 0.88)	1.88E-03	1.06 (0.87 - 1.28)	5.87E-01	0.87 (0.59 - 1.28)	4.77E-01

Peptide YY	P10082	1.34 (1.1 - 1.62)	3.36E-03	1.53 (1.27 - 1.86)	9.26E-06	1.43 (1.25 - 1.64)	1.77E-07
Periostin	Q15063	0.54 (0.43 - 0.68)	1.52E-07	0.75 (0.62 - 0.92)	5.50E-03	0.64 (0.47 - 0.88)	6.59E-03
Peroxiredoxin-1	Q06830	1.41 (1.15 - 1.72)	9.65E-04	1.01 (0.83 - 1.21)	9.51E-01	1.19 (0.85 - 1.65)	3.09E-01
Plasma protease C1 inhibitor	P05155	0.67 (0.53 - 0.84)	5.39E-04	0.76 (0.61 - 0.93)	9.66E-03	0.72 (0.61 - 0.83)	2.10E-05
Plasminogen activator inhibitor 1	P05121	1.46 (1.16 - 1.83)	1.19E-03	1.17 (0.97 - 1.42)	1.03E-01	1.29 (1.05 - 1.6)	1.71E-02
Protein S100-A9	P06702	1.53 (1.24 - 1.88)	6.16E-05				
Proto-oncogene tyrosine-protein kinase receptor Ret	P07949	1.39 (1.13 - 1.7)	1.75E-03	1.26 (1.02 - 1.56)	3.04E-02	1.33 (1.14 - 1.54)	1.73E-04
Pulmonary surfactant-associated protein D	P35247	1.43 (1.16 - 1.77)	6.82E-04	0.94 (0.77 - 1.12)	4.96E-01	1.16 (0.76 - 1.75)	4.96E-01
Renin	P00797	1.61 (1.32 - 1.99)	5.48E-06	1.45 (1.21 - 1.74)	5.16E-05	1.52 (1.33 - 1.74)	1.54E-09
Retinol-binding protein 4	P02753	1.3 (1.08 - 1.55)	3.68E-03				
RGM domain family member B	Q6NW40	0.64 (0.49 - 0.81)	3.52E-04	0.73 (0.59 - 0.9)	3.25E-03	0.69 (0.59 - 0.81)	5.33E-06
Sex hormone-binding globulin	P04278	0.62 (0.47 - 0.8)	2.41E-04	0.63 (0.51 - 0.77)	1.12E-05	0.62 (0.53 - 0.73)	1.04E-08
SLIT and NTRK-like protein 5	O94991	0.6 (0.47 - 0.76)	3.81E-05	0.78 (0.66 - 0.93)	4.38E-03	0.7 (0.54 - 0.9)	5.11E-03
SPARC-like protein 1	Q14515	0.69 (0.55 - 0.86)	1.22E-03	0.99 (0.82 - 1.21)	9.37E-01	0.83 (0.58 - 1.19)	3.10E-01
Transforming growth factor beta receptor type 3	Q03167	0.58 (0.46 - 0.73)	4.45E-06	0.74 (0.61 - 0.89)	1.40E-03	0.66 (0.52 - 0.85)	9.40E-04
Trypsin-1	P07477	0.63 (0.5 - 0.78)	4.06E-05	0.7 (0.58 - 0.84)	2.19E-04	0.67 (0.58 - 0.77)	4.35E-08
Tumor necrosis factor receptor superfamily member 11B	O00300	0.7 (0.58 - 0.86)	5.86E-04				
Tumor necrosis factor receptor superfamily member 13C	Q96RJ3	1.33 (1.12 - 1.57)	7.26E-04	0.88 (0.69 - 1.07)	2.43E-01	1.09 (0.72 - 1.63)	6.90E-01
Tumor necrosis factor-inducible gene 6 protein	P98066	0.58 (0.45 - 0.74)	2.27E-05	0.7 (0.57 - 0.85)	3.96E-04	0.65 (0.54 - 0.77)	1.28E-06
Tyrosine-protein kinase JAK2	O60674	0.72 (0.58 - 0.89)	2.54E-03	0.9 (0.76 - 1.07)	1.71E-01	0.81 (0.66 - 1.01)	5.64E-02
Tyrosine-protein kinase transmembrane receptor ROR1	Q01973	0.65 (0.49 - 0.84)	1.26E-03	0.95 (0.77 - 1.15)	6.12E-01	0.79 (0.54 - 1.15)	2.23E-01
WAP, Kazal, immunoglobulin, Kunitz and NTR domain-containing protein 2	Q8TEU8	0.69 (0.54 - 0.87)	2.10E-03	0.85 (0.7 - 1.04)	1.13E-01	0.77 (0.63 - 0.95)	1.31E-02
Wnt inhibitory factor 1	Q9Y5W5	0.5 (0.37 - 0.66)	2.97E-06	0.65 (0.54 - 0.79)	2.21E-05	0.58 (0.45 - 0.76)	6.44E-05

Table S3: KORA FDR significant results of Incident type 2 diabetes in KORA, HUNT as well as their combined effect using a meta-analysis random effects model. Model was adjusted for age and sex, body mass index (BMI), smoking and current hypertension status.

Protein Full Name	UniProt ID	KORA (N = 881)		HUNT (N = 794)		Combined	
		OR (95% CI)	p value	OR (95% CI)	p value	OR (95% CI)	Pvalue
Adiponectin	Q15848	0.41 (0.29 - 0.57)	1.84E-07				
Afamin	P43652	2.18 (1.64 - 2.94)	1.31E-07	1.09 (0.86 - 1.4)	4.90E-01	1.53 (0.78 - 3.03)	2.18E-01
Aminoacylase-1	Q03154	1.78 (1.34 - 2.37)	7.15E-05	1.6 (1.26 - 2.04)	1.27E-04	1.67 (1.39 - 2.01)	3.97E-08
Growth hormone receptor	P10912	1.74 (1.31 - 2.38)	2.43E-04	1.42 (1.07 - 1.88)	1.37E-02	1.56 (1.28 - 1.91)	1.64E-05
Insulin-like growth factor-binding protein 2	P18065	0.47 (0.34 - 0.65)	6.07E-06	0.57 (0.42 - 0.77)	2.91E-04	0.52 (0.42 - 0.65)	1.01E-08
Netrin receptor UNC5D	Q6UXZ4	0.57 (0.41 - 0.78)	4.37E-04	0.89 (0.71 - 1.12)	2.88E-01	0.72 (0.46 - 1.11)	1.39E-01
Pappalysin-1	Q13219	0.59 (0.45 - 0.77)	1.02E-04	1.02 (0.8 - 1.3)	8.45E-01	0.78 (0.45 - 1.34)	3.67E-01
Scavenger receptor cysteine-rich type 1 protein M130	Q86VB7	1.56 (1.23 - 2.01)	3.68E-04	0.98 (0.79 - 1.22)	8.53E-01	1.23 (0.78 - 1.95)	3.70E-01
Tartrate-resistant acid phosphatase type 5	P13686	1.74 (1.31 - 2.34)	1.97E-04	1.13 (0.9 - 1.44)	3.14E-01	1.39 (0.91 - 2.12)	1.27E-01
Transforming growth factor beta receptor type 3	Q03167	0.58 (0.44 - 0.77)	1.38E-04	0.83 (0.68 - 1.02)	5.55E-02	0.7 (0.5 - 0.99)	4.35E-02

Table S4: Overlap between our replicated proteins with type 2 diabetes associated genes from Xue et al, and Human Diabetes Proteome Project (HDPP) lists: 1000 diabetes related proteins list, islet of Langerhans proteome, rodent beta cell proteome and blood glycated proteins database.

Full Name	UniProt ID	Entrez Gene	Outcome in our study	T2D genes	HDPP lists				Any
					T2D 1000 proteins	Islet proteome	Rodent beta cell proteome	Glycated proteins	
Aminoacylase-1	Q03154	ACY1	Both			Y			Y
Growth hormone receptor	P10912	GHR	Incident		Y				Y
Insulin-like growth factor-binding protein 2	P18065	IGFBP2	Incident		Y	Y			Y
Alpha-L-iduronidase	P35475	IDUA	Prevalent			Y			Y
Apolipoprotein B	P04114	APOB	Prevalent		Y	Y			Y
Cathepsin Z	Q9UBR2	CTSZ	Prevalent			Y			Y
Cerebral dopamine neurotrophic factor	Q49AH0	CDNF	Prevalent						
Complement C2	P06681	C2	Prevalent	Y					Y
Galectin-3-binding protein	Q08380	LGALS3BP	Prevalent			Y			Y
Gelsolin	P06396	GSN	Prevalent			Y			Y
Hepatocyte growth factor receptor	P08581	MET	Prevalent			Y			Y
Kallikrein-7	P49862	KLK7	Prevalent						
Lysosomal protective protein	P10619	CTSA	Prevalent			Y			Y
Matrilin-2	O00339	MATN2	Prevalent						
Osteomodulin	Q99983	OMD	Prevalent						
Peptide YY	P10082	PYY	Prevalent		Y				Y
Periostin	Q15063	POSTN	Prevalent			Y			Y
Plasma protease C1 inhibitor	P05155	SERPING1	Prevalent			Y			Y
Renin	P00797	REN	Prevalent		Y				Y
RGM domain family member B	Q6NW40	RGMB	Prevalent						
Sex hormone-binding globulin	P04278	SHBG	Prevalent		Y				Y
SLIT and NTRK-like protein 5	O94991	SLITRK5	Prevalent						

Transforming growth factor beta receptor type 3	Q03167	TGFBR3	Prevalent						
Trypsin-1	P07477	PRSS1	Prevalent		Y	Y			Y
Tumor necrosis factor-inducible gene 6 protein	P98066	TNFAIP6	Prevalent			Y			Y
Wnt inhibitory factor 1	Q9Y5W5	WIF1	Prevalent						

Table S5: Results of the data analytics of replicated proteins:

(Excel sheet)

Table S6: Comparison between original model and drug-adjusted model with replicated prevalent type 2 diabetes proteins showing those that lost significance or showed different direction of effect estimate in any of the drug-adjusted models:

Drug (number of participants taking the drug)	C2		PYY		CATZ		N
	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value	
Original Result	2.01 (1.37 - 3.04)	6.63E-04	1.34 (1.1 - 1.62)	3.36E-03	1.41 (1.13 - 1.77)	2.27E-03	993
Antidiabetic Medications (52)	1.33 (0.9 - 2.18)	2.21E-01	0.9 (0.67 - 1.19)	4.71E-01	1.15 (0.86 - 1.54)	3.30E-01	993
Insulin (11)	1.93 (1.3 - 2.97)	1.81E-03	1.33 (1.08 - 1.63)	5.26E-03	1.39 (1.1 - 1.75)	5.32E-03	993
Oral antidiabetics (49)	1.34 (0.91 - 2.17)	2.02E-01	0.89 (0.66 - 1.17)	3.97E-01	1.19 (0.89 - 1.58)	2.31E-01	993
Oral antidiabetics without Metformin (27)	1.61 (1.08 - 2.52)	2.78E-02	1.14 (0.9 - 1.42)	2.63E-01	1.25 (0.97 - 1.6)	8.26E-02	993
Metformin (40)	1.37 (0.94 - 2.17)	1.44E-01	0.93 (0.71 - 1.21)	6.07E-01	1.27 (0.98 - 1.66)	7.51E-02	991

Table S7: Mendelian randomization results of the direction with type 2 diabetes as the exposure and the individual proteins as the outcome.

Protein	IVW Beta (SE)	IVW p-value	n SNPs	Pleiotropy test *	Protein summary statistics origin
Aminoacylase-1	-0.02 (0.07)	7.51E-01	96	2.01E-01	Suhre K
Growth hormone receptor	0.02 (0.04)	6.86E-01	120	5.60E-01	Sun BB
Insulin-like growth factor-binding protein 2	-0.05 (0.05)	2.47E-01	120	6.97E-01	Sun BB
Apolipoprotein B	-0.07 (0.04)	1.12E-01	120	3.24E-01	Sun BB
Cerebral dopamine neurotrophic factor	0.02 (0.04)	5.54E-01	120	8.09E-01	Sun BB
Plasma protease C1 inhibitor	-0.05 (0.04)	2.35E-01	120	7.84E-01	Sun BB
Complement C2	-0.02 (0.08)	8.35E-01	96	1.65E-01	Suhre K
Lysosomal protective protein	-0.003 (0.04)	9.39E-01	120	6.42E-01	Sun BB
Cathepsin Z	0.13 (0.04)	2.00E-03	120	1.80E-01	Sun BB
Gelsolin	0.03 (0.08)	6.61E-01	96	3.21E-01	Suhre K
Alpha-L-iduronidase	0.02 (0.04)	6.82E-01	120	6.87E-01	Sun BB
Kallikrein-7	-0.04 (0.04)	2.29E-01	120	9.56E-01	Sun BB
Galectin-3-binding protein	0.13 (0.07)	8.38E-02	96	9.82E-01	Suhre K
Matrilin-2	0.03 (0.04)	5.14E-01	120	1.24E-01	Sun BB
Hepatocyte growth factor receptor	-0.09 (0.05)	1.11E-01	120	4.38E-01	Sun BB
Osteomodulin	-0.02 (0.04)	5.70E-01	120	2.20E-01	Sun BB
Periostin	0.01 (0.04)	7.98E-01	120	7.68E-01	Sun BB
Peptide YY	0.05 (0.04)	1.70E-01	120	3.37E-01	Sun BB
Renin	0.08 (0.04)	3.15E-02	120	1.05E-01	Sun BB
RGM domain family member B	-0.01 (0.04)	8.52E-01	120	5.97E-01	Sun BB
Sex hormone-binding globulin	-0.01 (0.08)	8.55E-01	96	1.89E-02	Suhre K
SLIT and NTRK-like protein 5	0.01 (0.04)	7.49E-01	120	1.43E-02	Sun BB
Transforming growth factor beta receptor type 3	-0.01 (0.04)	8.30E-01	120	5.10E-01	Sun BB
Trypsin-1	-0.01 (0.04)	8.08E-01	120	6.16E-01	Sun BB
Tumor necrosis factor-inducible gene 6 protein	0.05 (0.04)	2.63E-01	120	5.41E-01	Sun BB
Wnt inhibitory factor 1	0.04 (0.04)	2.98E-01	120	1.72E-01	Sun BB

* Pleiotropy p value represents the p value of the intercept of Egger's regression.

Table S8: Mendelian randomization results of the direction with the proteins as the exposure and type 2 diabetes as the outcome. All causal effects were tested using Wald statistics.

Protein	Beta	SE	p value	SNP (IV)	IV Origin
Alpha-L-iduronidase	0.03	0.05	4.93E-01	rs76650972	Sun BB
Apolipoprotein B	0.05	0.03	1.16E-01	rs679899	Emilsson V
Cerebral dopamine neurotrophic factor	0.00	0.05	9.75E-01	rs11814337	Sun BB
Growth hormone receptor	-0.03	0.05	5.35E-01	rs34742008	Sun BB
Hepatocyte growth factor receptor	-0.01	0.05	8.07E-01	rs437	Sun BB
Kallikrein-7	0.01	0.05	7.74E-01	rs2691258	Sun BB
Matrilin-2	-0.06	0.05	2.77E-01	rs17831160	Sun BB
Peptide YY	0.03	0.05	5.84E-01	rs8074783	Emilsson V
Periostin	-0.04	0.03	1.91E-01	rs962462	Emilsson V
Plasma protease C1 inhibitor	0.00	0.01	7.62E-01	rs11229075	Sun BB
RGM domain family member B	0.02	0.04	5.65E-01	rs1563317	Sun BB
Sex hormone-binding globulin	-0.09	0.04	2.95E-02	rs858519	Emilsson V
Tumor necrosis factor-inducible gene 6 protein	-0.03	0.01	5.41E-02	rs289828	Sun BB

Table S9: Information on replicated proteins' validation adapted from Emilsson V. et al.:

Target Full Name	UniProt	Gene Symbol	DDA	MRM	Cis effect	Cis-Trans effects	Protein module
Matrilin-2	O00339	MATN2	MATN2				PM26
SLIT and NTRK-like protein 5	O94991	SLITRK5			SLITRK5		PM26
Renin	P00797	REN	REN	REN			No module
Apolipoprotein B	P04114	APOB	APOB	APOB	APOB	APOB	PM11
Sex hormone-binding globulin	P04278	SHBG	SHBG		SHBG	SHBG	No module
Plasma protease C1 inhibitor	P05155	SERPING1	SERPING1		SERPING1	SERPING1	PM27
Gelsolin	P06396	GSN	GSN		GSN		PM27
Complement C2	P06681	C2	C2		C2	C2	No module
Trypsin-1	P07477	PRSS1					No module
Hepatocyte growth factor receptor	P08581	MET		MET	MET	MET	PM27
Peptide YY	P10082	PYY			PYY		PM26
Lysosomal protective protein	P10619	CTSA	CTSA			CTSA	No module
Growth hormone receptor	P10912	GHR			GHR		No module
Insulin-like growth factor-binding protein 2	P18065	IGFBP2	IGFBP2				PM10
Alpha-L-iduronidase	P35475	IDUA		IDUA	IDUA		No module
Kallikrein-7	P49862	KLK7	KLK7		KLK7		PM27
Tumor necrosis factor-inducible gene 6 protein	P98066	TNFAIP6			TNFAIP6		PM27
Aminoacylase-1	Q03154	ACY1	ACY1	ACY1			PM23
Transforming growth factor beta receptor type 3	Q03167	TGFBR3					PM26
Galectin-3-binding protein	Q08380	LGALS3BP			LGALS3BP		PM6
Periostin	Q15063	POSTN			POSTN		PM13
Cerebral dopamine neurotrophic factor	Q49AH0	CDNF			CDNF		PM2
RGM domain family member B	Q6NW40	RGMB					PM26
Osteomodulin	Q99983	OMD					PM27
Cathepsin Z	Q9UBR2	CTSZ			CTSZ		PM26
Wnt inhibitory factor 1	Q9Y5W5	WIF1				WIF1	PM13

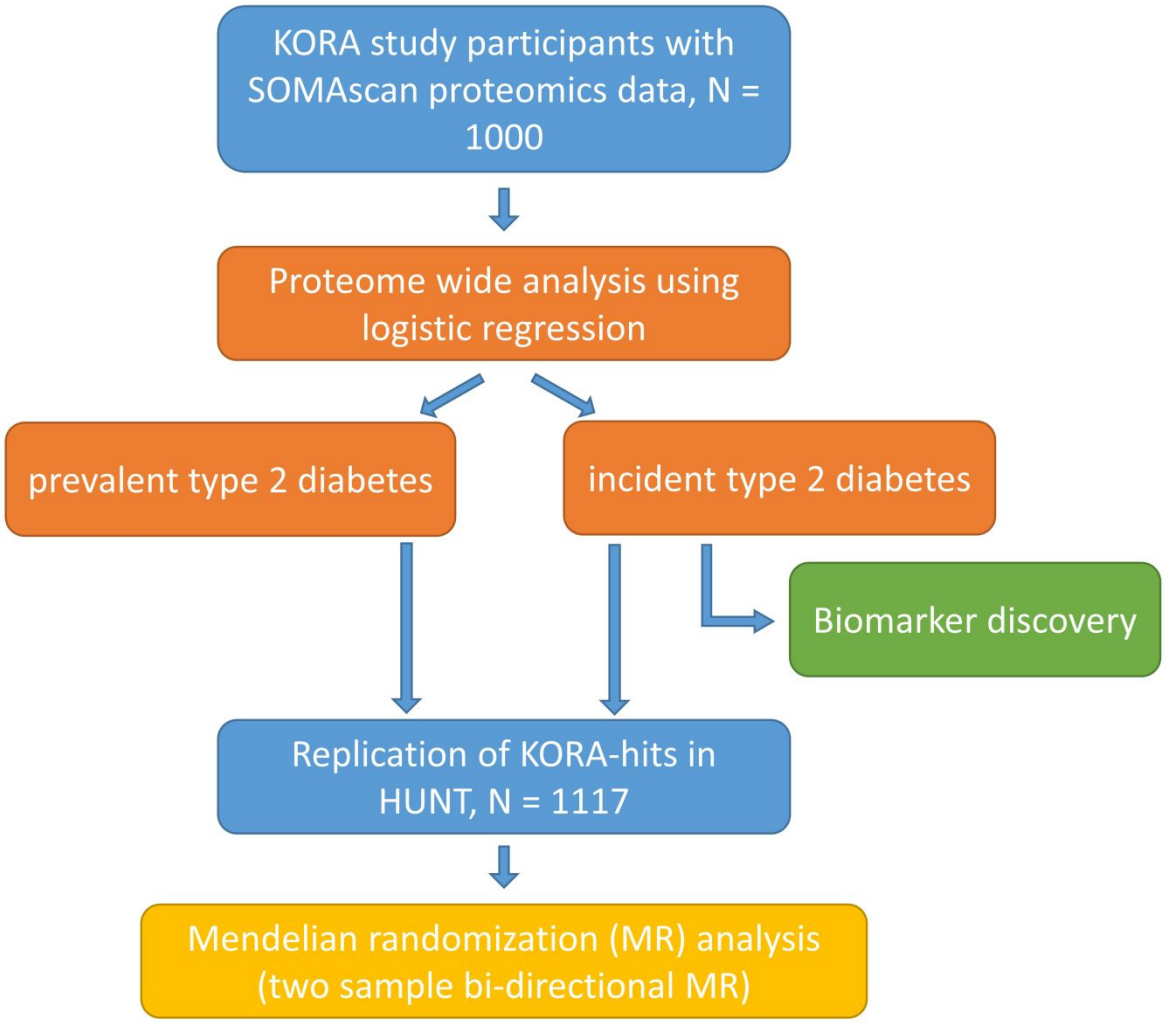

Figure S1: Analysis flowchart.

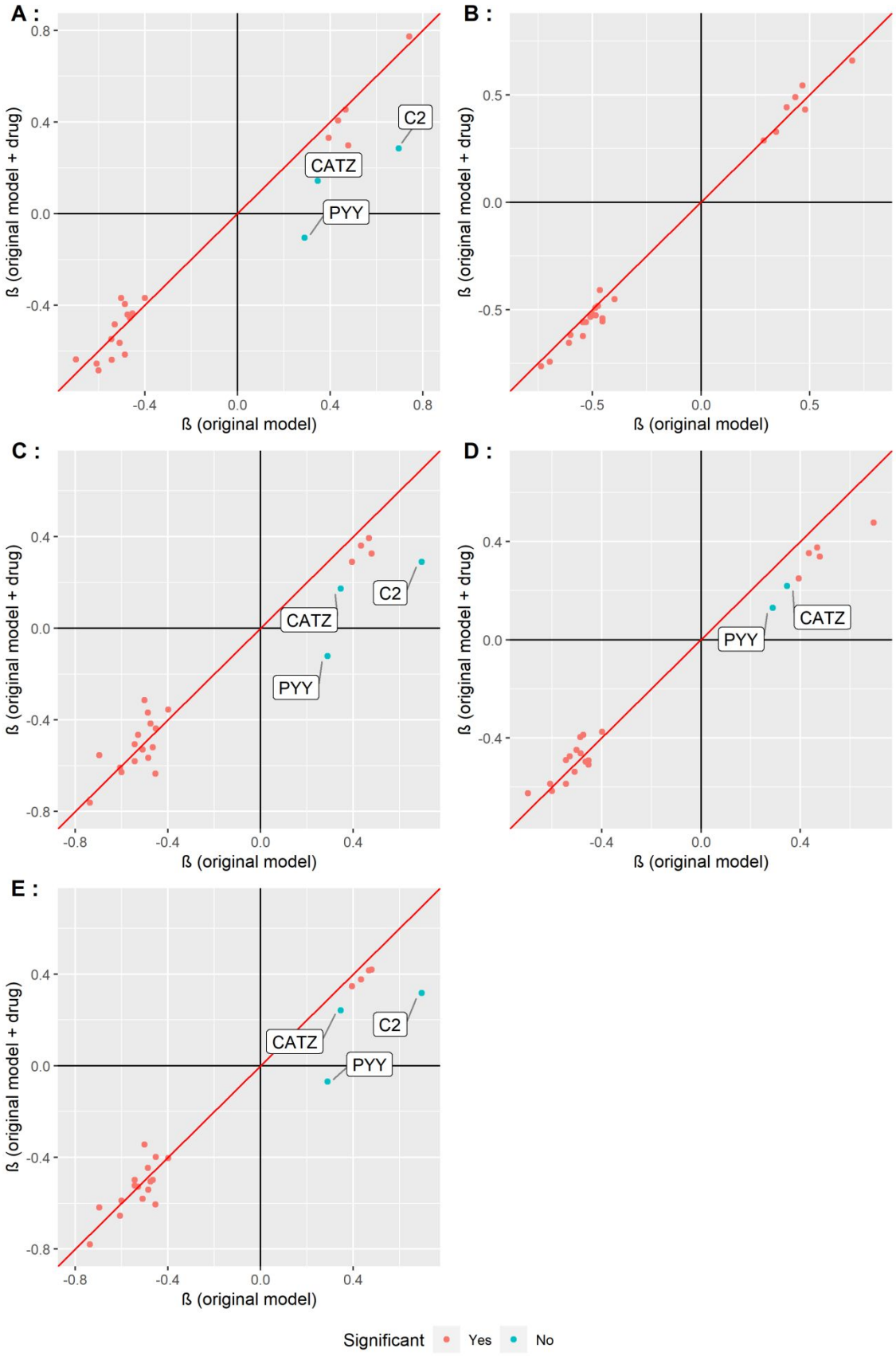

Figure S2: Results of drug-adjusted model of replicated prevalent type 2 diabetes associations. (A: Antidiabetic Medications, B: Insulin, C: Oral antidiabetics, D: Oral antidiabetics without Metformin, E: Metformin).

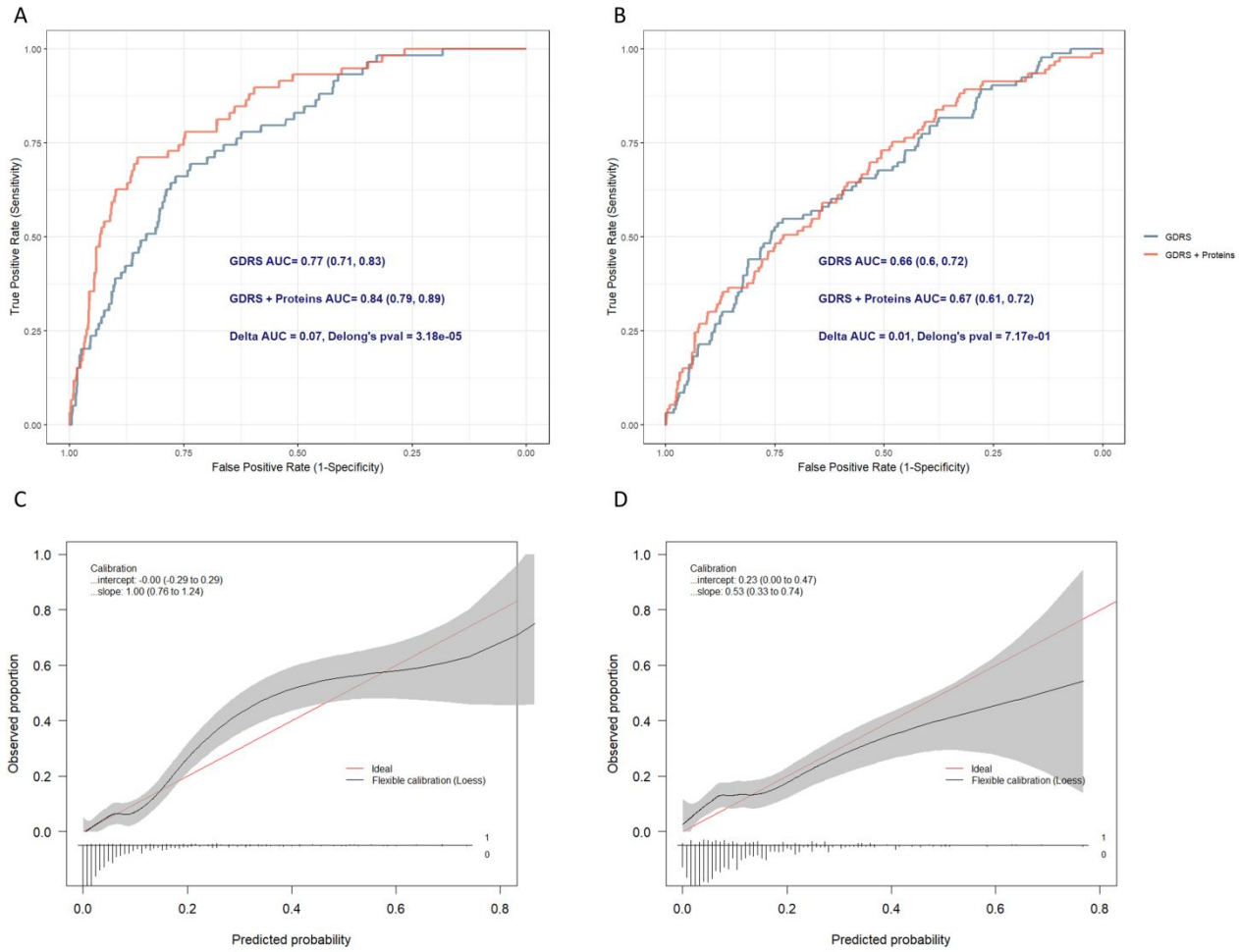

Figure S3: Performance of LASSO selected biomarkers on top of GDRS compared to GDRS in KORA and HUNT: assessing discrimination using receiver operating characteristic curve and c-statistic (A: KORA, B: HUNT), and assessing calibration (C: KORA, D: HUNT).

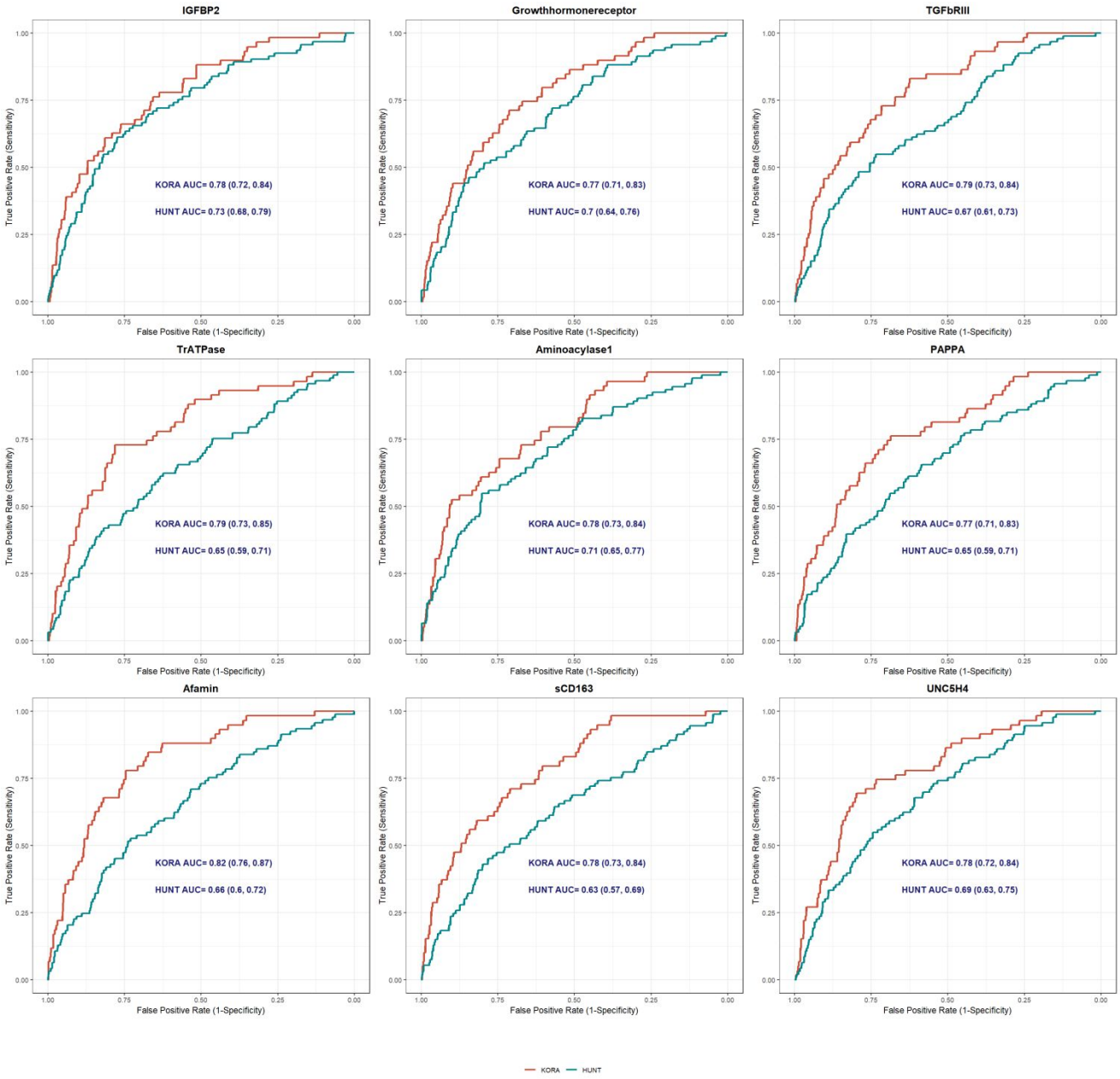

Figure S4: Performance of individual proteins on top of GDRS in both KORA and HUNT.

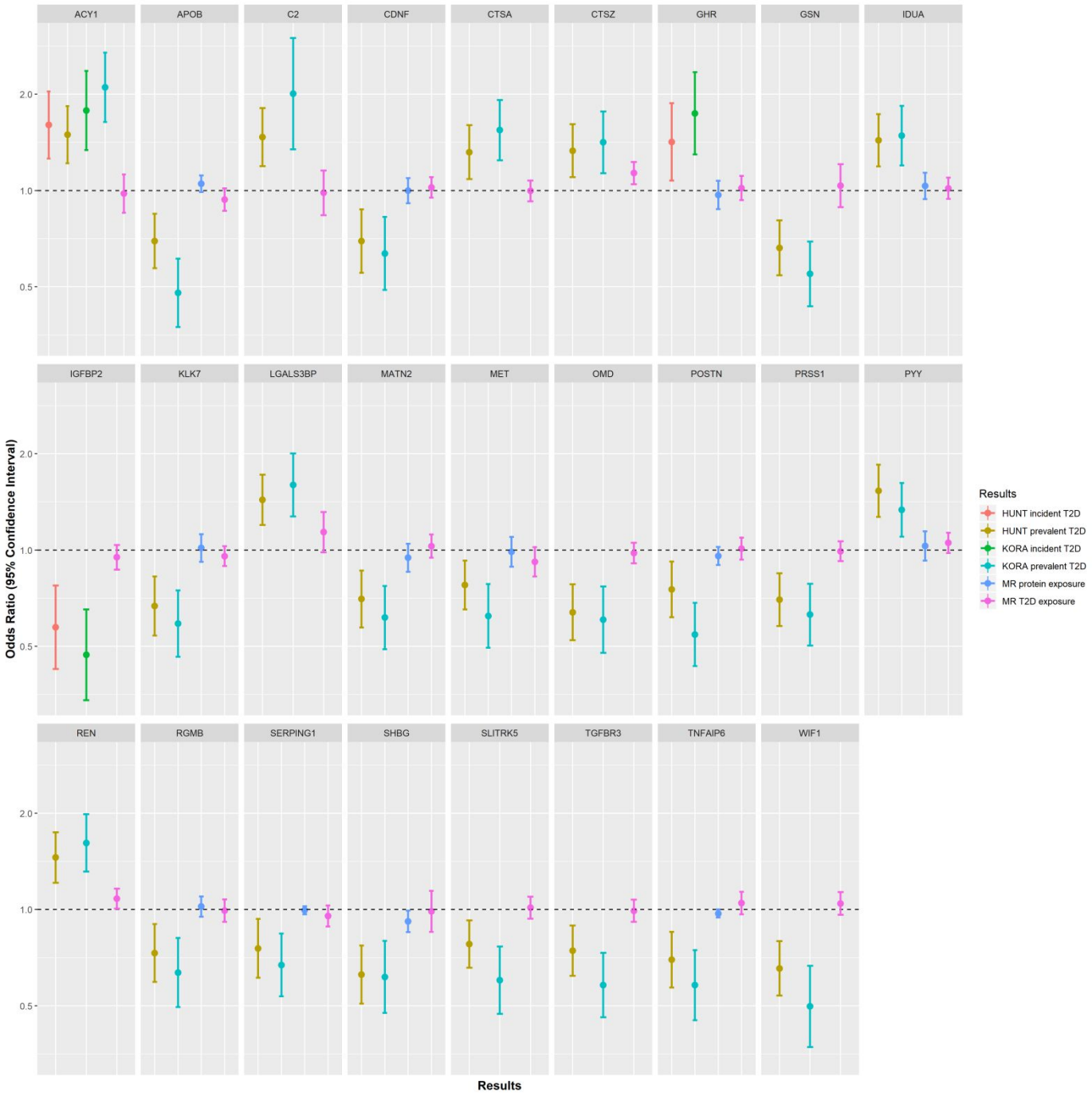

Figure S5: Forest plot of proteome wide analysis in KORA and HUNT as well as Mendelian randomization results for replicated proteins. Odds ratio are per SD change of protein level.

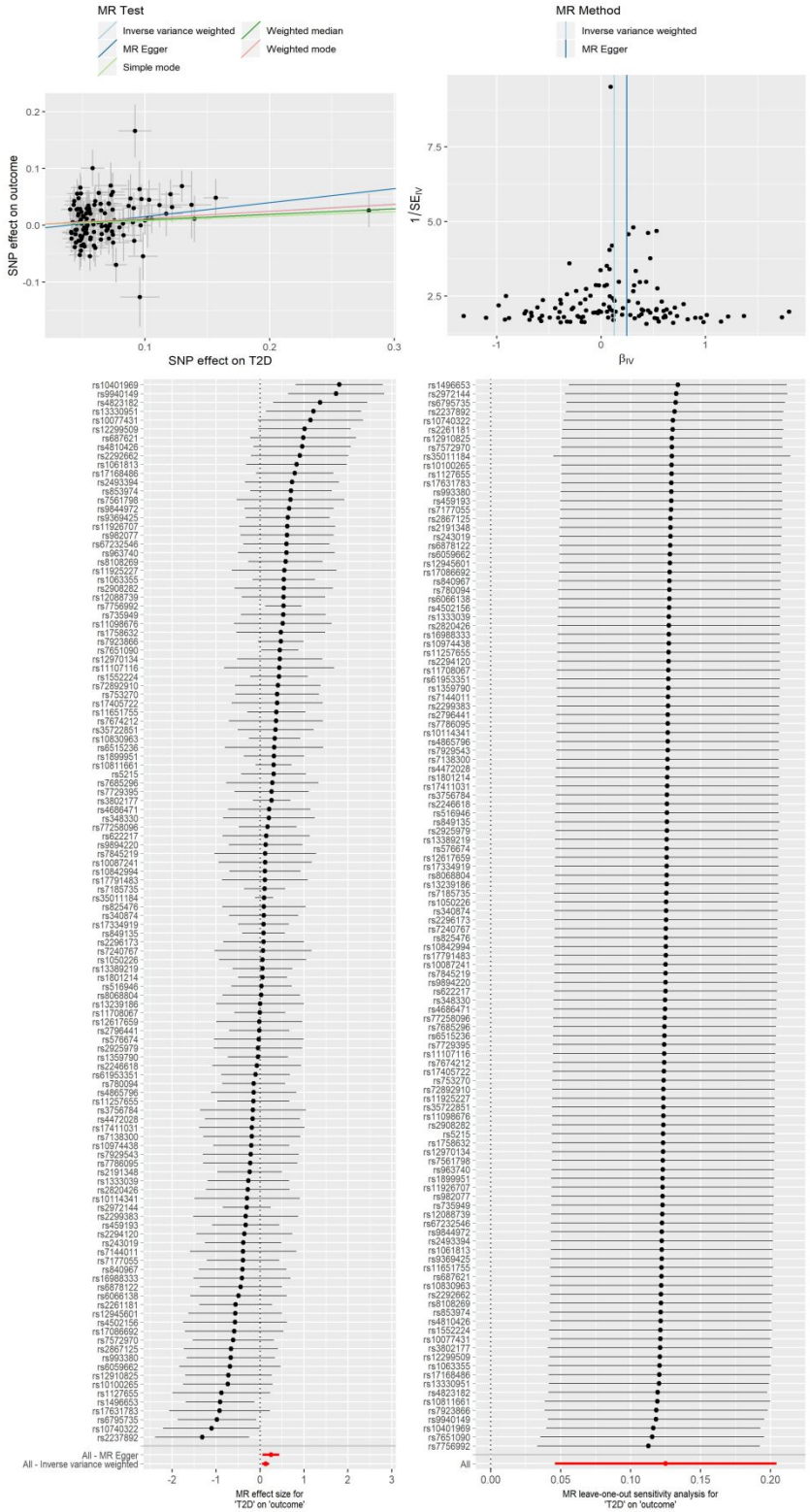

Figure S6: Sensitivity analyses of the Mendelian randomization analysis with cathepsin Z as the outcome. From top left to right: a: Scatter plot of the IVs' effects on exposure and outcome with causal analysis fitted lines plotted to check if there are any outliers driving the results; b: Funnel plot of the IVs' effects on the exposure to check for pleiotropy; c: Forest plot of the casual effect of each SNP to check for any potential outliers driving the results; d: Leave-one-out analysis to check if the results are dependent on any specific IV.

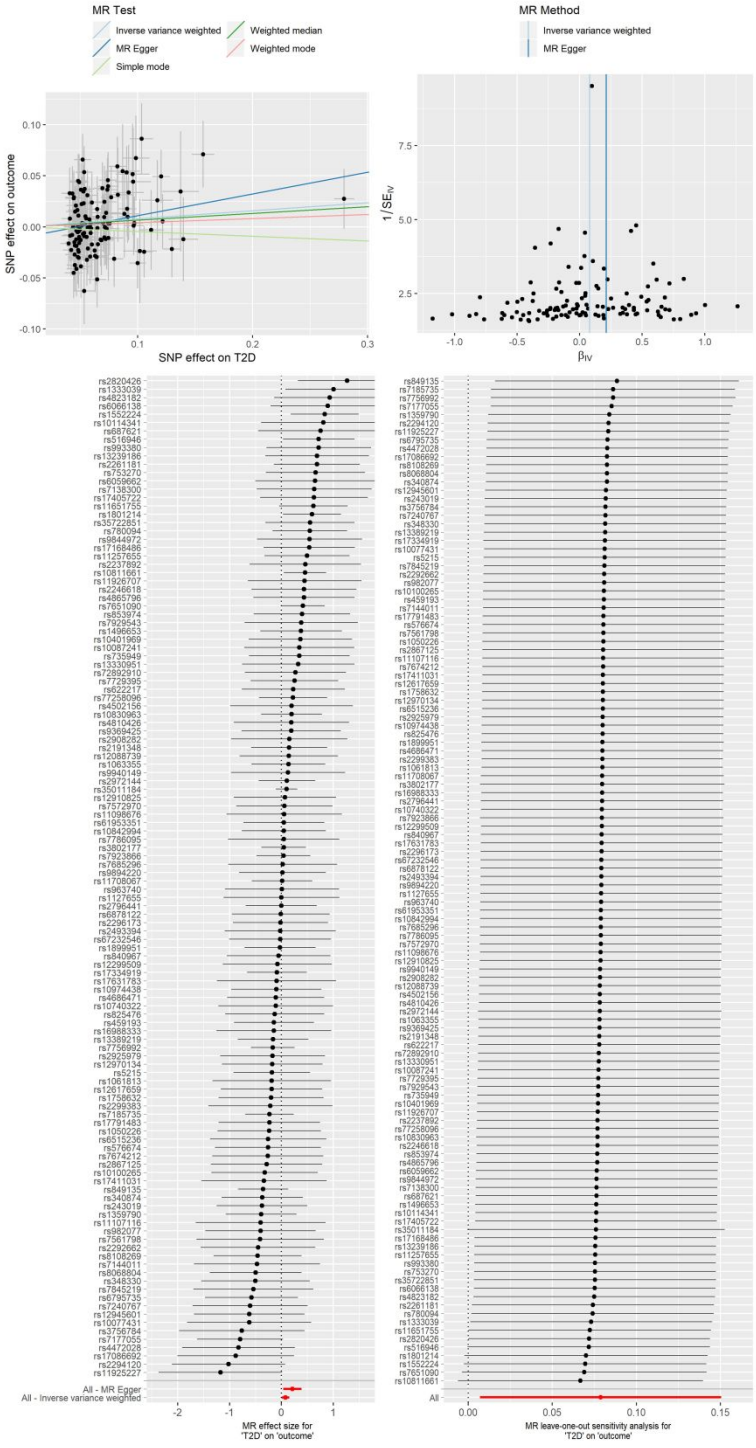

Figure S7: Sensitivity analyses of the Mendelian randomization analysis with renin as the outcome. From top left to right: a: Scatter plot of the IVs' effects on exposure and outcome with causal analysis fitted lines plotted to check if there are any outliers driving the results; b: Funnel plot of the IVs' effects on the exposure to check for pleiotropy; c: Forest plot of the casual effect of each SNP to check for any potential outliers driving the results; d: Leave-one-out analysis to check if the results are dependent on any specific IV.

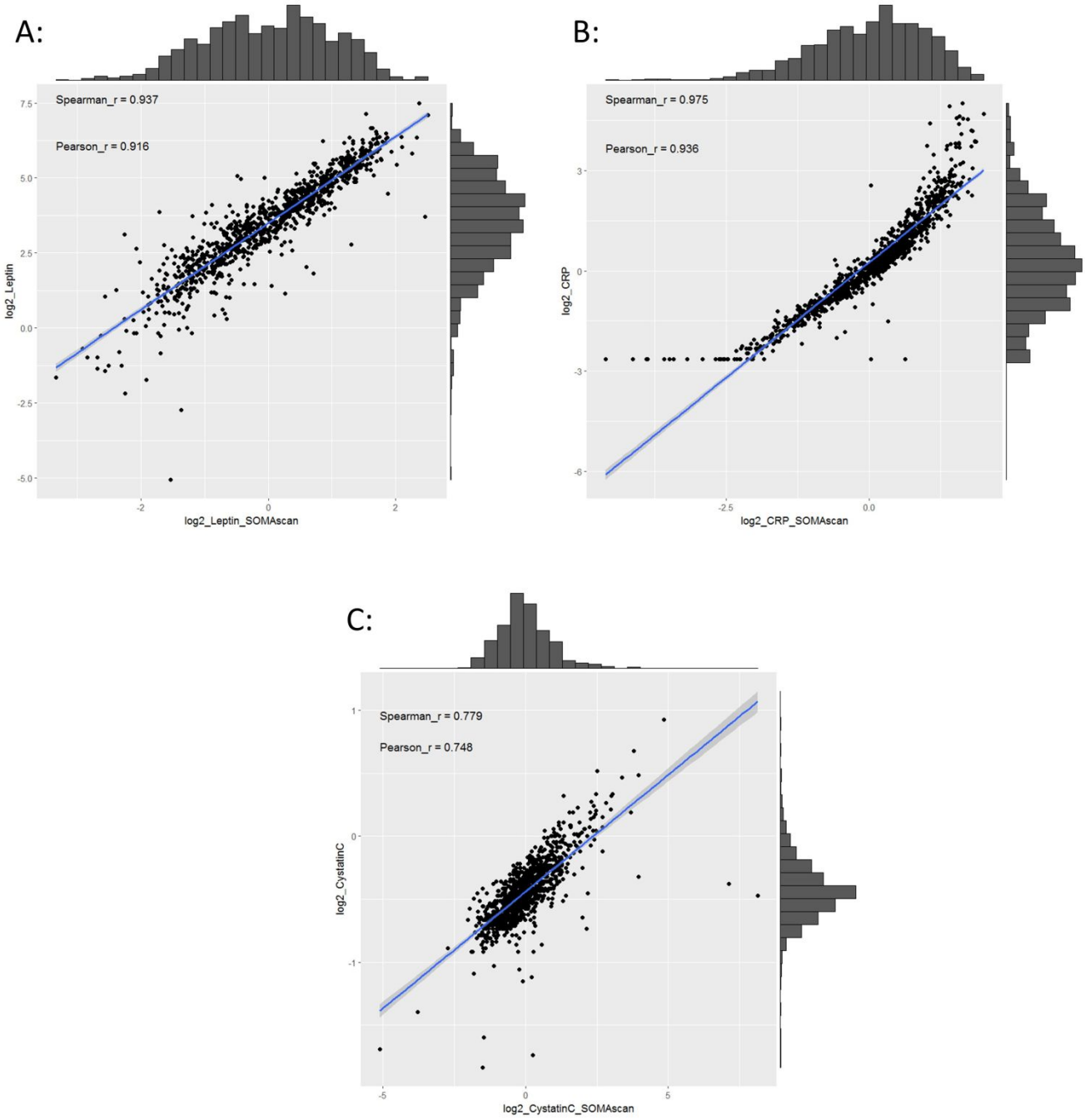

Figure S8: Correlation between SOMAscan measured and otherwise measured biomarkers in KORA (A: leptin by ELISA; B: CRP by nephelometry; C: cystatin-C by nephelometry.

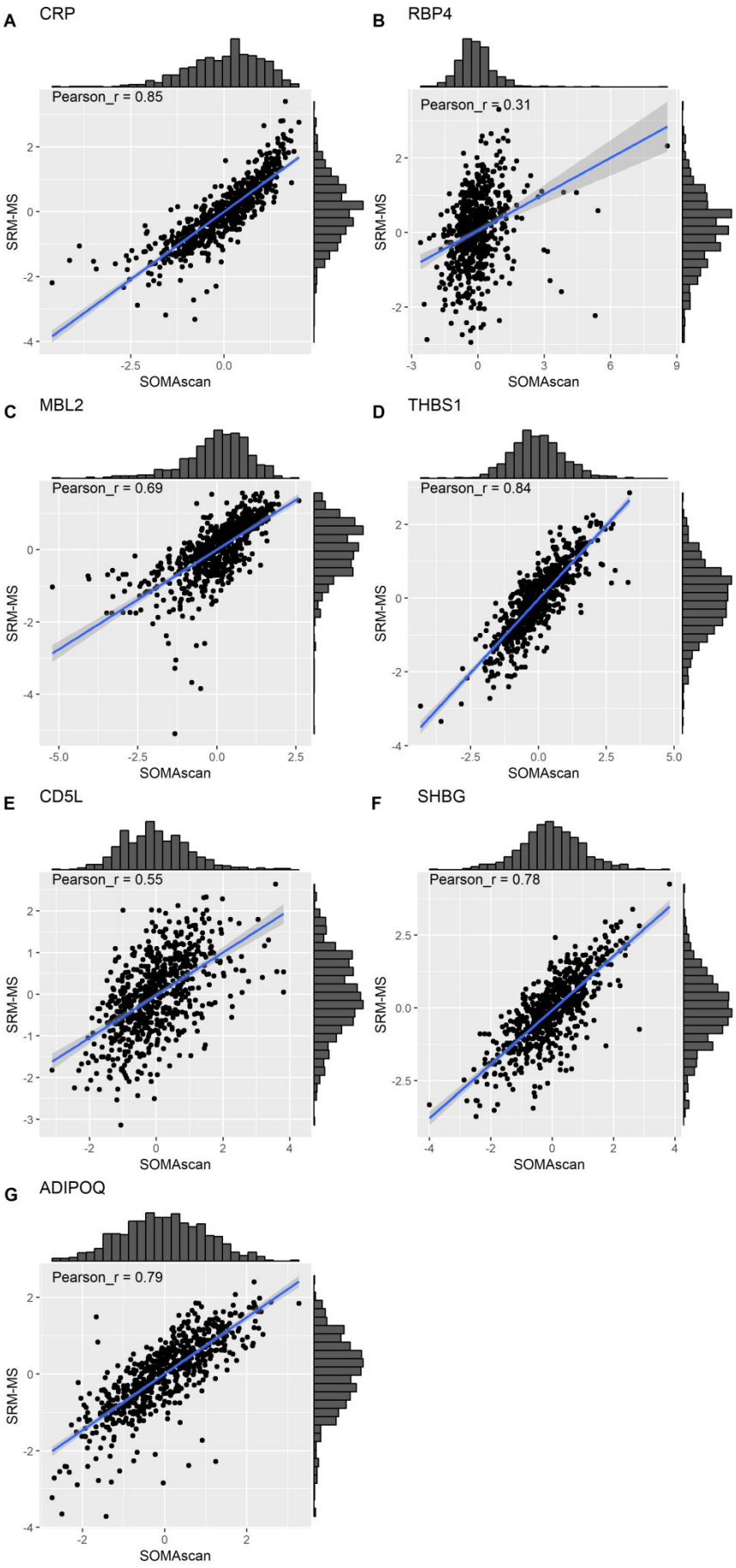

Figure S9: Correlation between SOMAscan measured and SRM-MS measured biomarkers in KORA

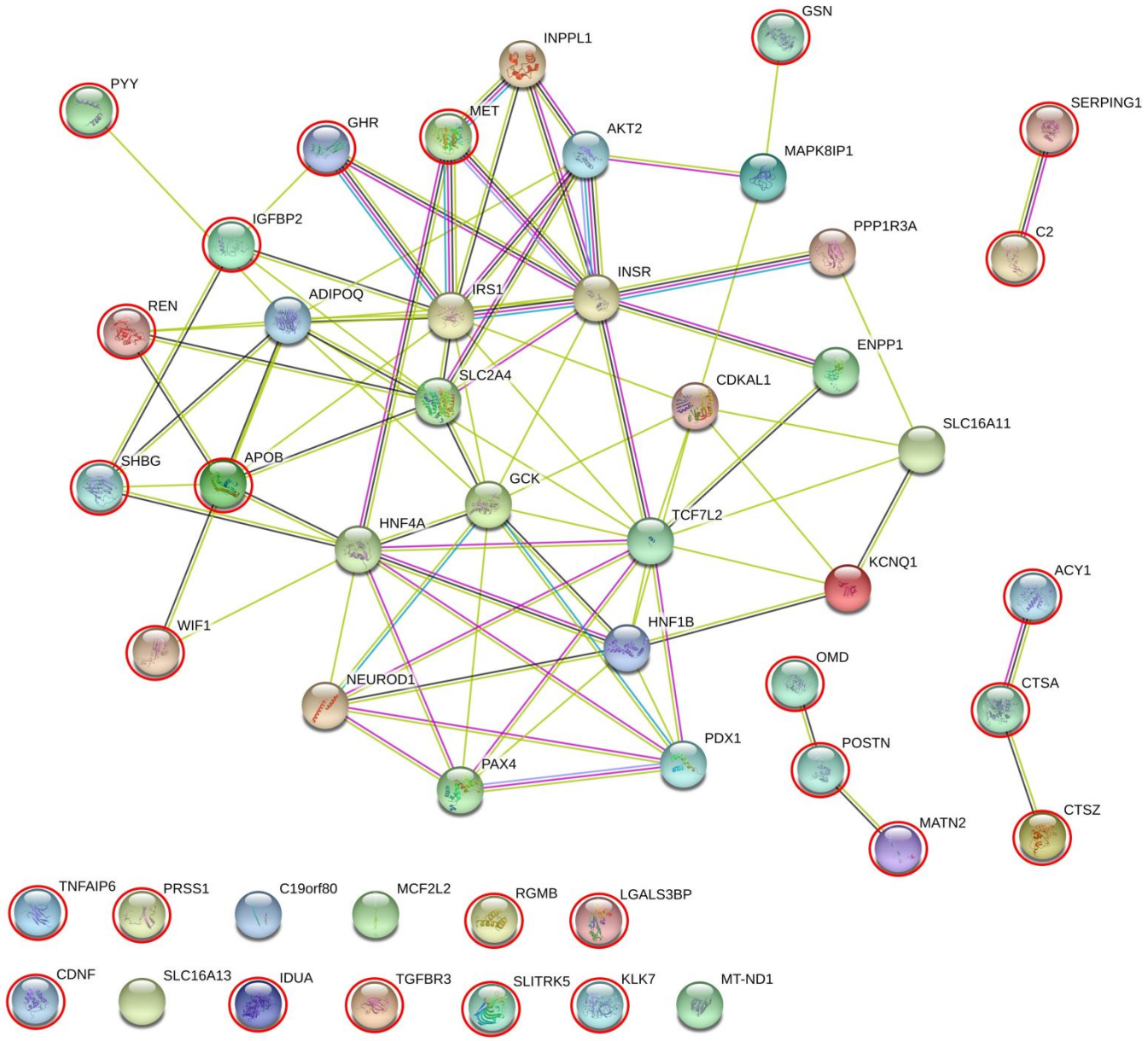

Figure S10: Network analysis using STRING featuring our replicated proteins, which are marked by red circles, and UniProt reported proteins associated with type 2 diabetes (https://version-11-0.string-db.org/cgi/network.pl?networkId=OzpbBXbgS2PW). Of the UniProt curated proteins, only two proteins were measured by SOMAscan. Adiponectin was significant in KORA but failed quality control in HUNT and could not be replicated and insulin receptor was not significant in our results.

References

1. Mühlenbruch K, Joost H-G, Boeing H, Schulze M. Risk prediction for type 2 diabetes in the German population with the updated German Diabetes Risk Score (GDRS). Ernährungs Umschau 2014;61:90-93
2. Paprott R, Mühlenbruch K, Mensink GBM, Thiele S, Schulze MB, Scheidt-Nave C, Heidemann C. Validation of the German Diabetes Risk Score among the general adult population: findings from the German Health Interview and Examination Surveys. BMJ Open Diabetes Research & Care 2016;4:e000280
3. Perna L, Mielck A, Lacruz ME, Emeny RT, von Eisenhart Rothe A, Meisinger C, Ladwig K-H. The association between resilience and diabetic neuropathy by socioeconomic position: Cross-sectional findings from the KORA-Age study. Journal of Health Psychology 2013;20:1222-1228
4. Emilsson V, Ilkov M, Lamb JR, Finkel N, Gudmundsson EF, Pitts R, Hoover H, Gudmundsdottir V, Horman SR, Aspelund T, Shu L, Trifonov V, Sigurdsson S, Manolescu A, Zhu J, Olafsson O, Jakobsdottir J, Lesley SA, To J, Zhang J, Harris TB, Launer LJ, Zhang B, Eiriksdottir G, Yang X, Orth AP, Jennings LL, Gudnason V. Co-regulatory networks of human serum proteins link genetics to disease. Science (New York, NY) 2018;361:769-773
5. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic acids research 2019;47:D607-d613
6. Consortium TU. UniProt: a worldwide hub of protein knowledge. Nucleic acids research 2019;47:D506-D515

Name
UniProtID
Name
Protein origin
Gene
Network
TDL
DTO.Class
Panther
Subcell_Location
GO.Experimental.MF/BP.Leaf.Terms
OMIM.Phenotype(s)
JAX/MGI.Human.Ortholog.Phenotype(s)
IMPC.Ortholog.Phenotype(s)
GWAS.Phenotype(s)
Pathways
DISEASES

Description
protin UniProt ID
full protein name
protein origin, which could be our replicated results or type 2 diabetes associated proteins extracted from Uni
gene name of the protein
whether the protein is part of the STRING netwok connecting replicated proteins to UniProt reported proteins
targer development level infomration from PHAROS
Drug Target Ontology category as per http://drugtargetontology.org/ and paper https://doi.org/10.1186/s133
protein analysis through evolutionary relationships is a protein classification scheme https://en.wikipedia.org/
subcellar location of the protein
Gene Ontology Molecular Function (MF) or Biological Process (BP) Leaf term, aka GO terms for which no child
data from OMIM on phenotypes related to that protein
data from JAX/MGI on mouse phenotype related to that protein
data from IMPC on mouse phenotype related to that protein
data from GWAS catalog
pathways obtained from KEGG and several other resources (Reactome, PathwayCommons, WikiPathways, Un
disease association obtained from the diseases platform

Supplementary Table S5: Results of the data analytics of replicated proteins

UniProtID	Name	Protein origin	Gene
P35475	Alpha-L-iduronidase	Prevalent T2D	IDUA
Q03154	Aminoacylase-1	Prevalent and incident T2D	ACY1
P10619	Lysosomal protective protein	Prevalent T2D	CTSA
Q9UBR2	Cathepsin Z	Prevalent T2D	CTSZ
P04114	Apolipoprotein B-100	Prevalent T2D	APOB
Q49AH0	Cerebral dopamine neurotrophic factor	Prevalent T2D	CDNF
P05155	Plasma protease C1 inhibitor	Prevalent T2D	SERPING1
P06681	Complement C2	Prevalent T2D	C2
Q08380	Galectin-3-binding protein	Prevalent T2D	LGALS3BP
P06396	Gelsolin	Prevalent T2D	GSN
P08581	Hepatocyte growth factor receptor	Prevalent T2D	MET
P49862	Kallikrein-7	Prevalent T2D	KLK7
O00339	Matrilin-2	Prevalent T2D	MATN2
Q99983	Osteomodulin	Prevalent T2D	OMD
Q15063	Periostin	Prevalent T2D	POSTN
P10082	Peptide YY	Prevalent T2D	PYY
P00797	Renin	Prevalent T2D	REN
Q6NW40	RGM domain family member B	Prevalent T2D	RGMB
P04278	Sex hormone-binding globulin	Prevalent T2D	SHBG
O94991	SLIT and NTRK-like protein 5	Prevalent T2D	SLITRK5
Q03167	Transforming growth factor beta receptor type 3	Prevalent T2D	TGFBR3
P07477	Trypsin-1	Prevalent T2D	PRSS1
P98066	Tumor necrosis factor-inducible gene 6 protein	Prevalent T2D	TNFAIP6
Q9Y5W5	Wnt inhibitory factor 1	Prevalent T2D	WIF1
P10912	Growth hormone receptor	Incident T2D	GHR
P18065	Insulin-like growth factor-binding protein 2	Incident T2D	IGFBP2
P14672	Solute carrier family 2, facilitated glucose transp	UniProtKB	SLC2A4
Q9NQB0	Transcription factor 7-like 2 (HMG box transcrip	UniProtKB	TCF7L2
Q16821	Protein phosphatase 1 regulatory subunit 3A (P	UniProtKB	PPP1R3A
P52945	Pancreas/duodenum homeobox protein 1 (PDX-	UniProtKB	PDX1
Q7RTY0	Monocarboxylate transporter 13 (MCT 13) (Solu	UniProtKB	SLC16A13
Q8NCK7	Monocarboxylate transporter 11 (MCT 11) (Solu	UniProtKB	SLC16A11
O15357	Phosphatidylinositol 3,4,5-trisphosphate 5-phos	UniProtKB	INPPL1
Q13562	Neurogenic differentiation factor 1 (NeuroD) (N	UniProtKB	NEUROD1
O43316	Paired box protein Pax-4	UniProtKB	PAX4
P35568	Insulin receptor substrate 1 (IRS-1)	UniProtKB	IRS1
P35680	Hepatocyte nuclear factor 1-beta (HNF-1-beta) (	UniProtKB	HNF1B
P03886	NADH-ubiquinone oxidoreductase chain 1 (EC 7.	UniProtKB	MT-ND1
P41235	Hepatocyte nuclear factor 4-alpha (HNF-4-alpha	UniProtKB	HNF4A
P51787	Potassium voltage-gated channel subfamily KQT	UniProtKB	KCNQ1
P06213	Insulin receptor (IR) (EC 2.7.10.1) (CD antigen CI	UniProtKB	INSR
P35557	Hexokinase-4 (HK4) (EC 2.7.1.1) (Glucokinase) (H	UniProtKB	GCK
Q15848	Adiponectin (30 kDa adipocyte complement-rela	UniProtKB	ADIPOQ
P31751	RAC-beta serine/threonine-protein kinase (EC 2.	UniProtKB	AKT2
Q86YR7	Probable guanine nucleotide exchange factor M	UniProtKB	MCF2L2
Q6UXH0	Angiopoietin-like protein 8 (Betatrophin) (Lipasi	UniProtKB	ANGPTL8
Q5VV42	Threonylcarbamoyladenosine tRNA methylthiot	UniProtKB	CDKAL1
Q9UQF2	C-Jun-amino-terminal kinase-interacting protein	UniProtKB	MAPK8IP1
P22413	Ectonucleotide pyrophosphatase/phosphodiest	UniProtKB	ENPP1

Network	TDL	DTO Class	Panther	Subcell_Lo	GO Experimenta	OMIM Phenotype(s)
NotT2DNet	Tbio	Glycosidase	PC00110:gl	Lysosome |	GO:0003940 | F:L	MIM Number: 252800;
T2DNet	Tbio			Cytosol | Vesicles		MIM Number: 104620;
T2DNet	Tchem	Serine prot	PC00203:se	Vesicles	GO:1904715 | P:n	MIM Number: 613111;
T2DNet	Tchem	Cysteine pr	PC00191:pr	Vesicles	GO:0010757 | P:n	MIM Number: 603169
T2DNet	Tchem			Vesicles | C	GO:0008201 | F:h	MIM Number: 107730;
NotT2DNet	Tbio			Endoplasmic Reticulum		MIM Number: 611233
T2DNet	Tbio	Serine prot	PC00191:pr	Plasma me	GO:0004867 | F:s	MIM Number: 606860;
T2DNet	Tbio			Vesicles		MIM Number: 613927;
NotT2DNet	Tbio			Plasma membrane		MIM Number: 600626
T2DNet	Tbio	Non-motor	PC00041:ac	Actin filam	GO:1990000 | P:a	MIM Number: 137350;
T2DNet	Tclin	Kinase; MET family		Cytosol | Pl	GO:0001886 | P:e	MIM Number: 164860;
NotT2DNet	Tchem	Serine prot	PC00203:se	Plasma membrane		
T2DNet	Tbio	Receptor	PC00197:re	Secreted		MIM Number: 602108
T2DNet	Tbio			Golgi apparatus | Golgi lume		0
T2DNet	Tbio	Signaling	PC00069:ce	Golgi appar	GO:0071307 | P:c	MIM Number: 608777
T2DNet	Tbio	Neuropepti	PC00162:ne	Secreted		MIM Number: 600781
T2DNet	Tclin	Aspartic pr	PC00053:as	Plasma me	GO:0002003 | P:a	MIM Number: 179820;
T2DNet	Tbio			Nucleoplasm		MIM Number: 612687
T2DNet	Tchem			Vesicles		MIM Number: 182205
NotT2DNet	Tbio			Nucleoplasm | Golgi apparat		MIM Number: 609680
T2DNet	Tbio			Cytosol	GO:0050431 | F:tr	MIM Number: 600742
T2DNet	Tclin	Serine prot	PC00203:se	Vesicles		MIM Number: 276000;
T2DNet	Tbio			Secreted		MIM Number: 600410
T2DNet	Tbio	Calcium-bir	PC00060:ca	lcium-binding protein		MIM Number: 605186
T2DNet	Tclin	Cytokine	PC00090:de	Cytosol | Cy	GO:0042803 | F:p	MIM Number: 600946;
T2DNet	Tchem	Protease in	PC00191:pr	Endoplasm	GO:0031994 | F:ir	MIM Number: 146731
T2DNet	Tbio	Class I transporters		Plasma me	GO:0032869 | P:c	MIM Number: 138190
T2DNet	Tbio			Nucleoplas	GO:0070016 | F:a	MIM Number: 602228;
T2DNet	Tbio	Phosphatas	PC00184:pl	Membrane protein		MIM Number: 600917;
T2DNet	Tbio			Nucleoplas	GO:0003309 | P:t	MIM Number: 608769;
NotT2DNet	Tdark	SLC16 fami	PC00227:tr	Golgi apparatus		
T2DNet	Tbio	SLC16 fami	PC00227:tr	ER membrane | Nuclear men		MIM Number: 615765
T2DNet	Tchem			Golgi appar	GO:0042169 | F:S	MIM Number: 600829;
T2DNet	Tbio	Nuclease	PC00218:tr	Nucleoplas	GO:0070888 | F:E	MIM Number: 601724;
T2DNet	Tbio			Nucleoplasm | Nucleus		MIM Number: 167413;
T2DNet	Tbio			Cytosol | Nu	GO:0005158 | F:ir	MIM Number: 147545;
T2DNet	Tbio	Nucleic aci	PC00218:tr	Nucleoplas	GO:0042803 | F:p	MIM Number: 189907;
T2DNet	Tclin	Dehydroge	PC00092:de	Mitochondria | Mitochondrial membrane		
T2DNet	Tchem	2A. Hepato	PC00218:tr	Nucleoplas	GO:0042803 | F:p	MIM Number: 600281;
T2DNet	Tclin	Kv7.1/KCNQ1 sub-subf		Plasma me	GO:0005516 | F:c	MIM Number: 607542;
T2DNet	Tclin	INSR family		Vesicles | Pl	GO:0001540 | F:a	MIM Number: 147670;
T2DNet	Tchem	Hexokinase family		Golgi appar	GO:0005524 | F:A	MIM Number: 138079;
T2DNet	Tbio			Endoplasm	GO:0042803 | F:p	MIM Number: 605441;
T2DNet	Tchem	AKT family	PC00167:nc	Nucleus | V	GO:0005524 | F:A	MIM Number: 164731;
NotT2DNet	Tbio	Signaling	PC00207:si	Cytosol | Plasma membrane		
T2DNet	Tbio			Golgi apparatus | Nucleoplas		MIM Number: 616223
T2DNet	Tbio			Vesicles	GO:0035598 | F:N	MIM Number: 611259
T2DNet	Tbio	Signaling	PC00207:si	Plasma me	GO:0005078 | F:M	MIM Number: 604641;
T2DNet	Tchem	Nucleotide	PC00196:py	Plasma me	GO:0005524 | F:A	MIM Number: 173335;

JAX/MGI Human Or	IMPC Ortholog Phe	GWAS Phenotype(s)	Pathways	DISEASES
MP:0005367:renal/urinary sys_phen|M		Bone mineral density (h	KEGG:Lysosome|KEC	Mucopolysaccharido
; Phenotype: Aminoacylase 1 deficiency, 609924 (3)			KEGG:2-Oxocarboxyl	Orotic aciduria (ZSc
MP:0005387:immune sys_phen|MP:0	005367:renal/urinary sys_		KEGG:Renin-angioter	Lysosomal storage d
		Mean platelet volume (h	KEGG:Lysosome|KEGG:Apoptosis|RxT:As	
MP:0005385:c-v sys_phen|MP:000538		Educational attainment	KEGG:Fat digestion a	Lipid metabolism dis
		Parental extreme longevity (95 years and old		Parkinson's disease
MP:0005385:c-v sys_phen		General cognitive ability	KEGG:Complement a	Angioedema (ZScore
; Phenotype: C2 deficiency, 217000 (3);		General cognitive ability	KEGG:Complement a	Age related macular
MP:0005376:homeostasis/metabolism		phenotype|MP:000538	RxT:Hemostasis|RxT:	Liver disease (ZScor
MP:0010768:mortality/aging|MP:0005		Pulse pressure (http://w	KEGG:Fc gamma R-m	Amyloidosis (ZScore
MP:0005379:endocrine/exocrine gland		Triglycerides (http://ww	KEGG:PI3K-Akt signal	Cancer (ZScore: 6.36
		Blood protein levels (htt	RxT:Degradation of t	Netherton syndrom
		Adverse response to chemotherapy (neutrop		Achondrogenesis (Z
MP:0005390:skeleton phenotype|MP:(		Lung function (FVC) (htt	RxT:Defective B4GAL	Dystonia (ZScore: 5.
MP:0005387:immur	MP:0000063:decr l	Blood protein levels (htt	W_P:Amplification a	Asthma (ZScore: 5.4
MP:0005390:skeleton phenotype|MP:(		Diverticular disease (htt	KEGG:Neuroactive lig	Obesity (ZScore: 6.1
MP:0010768:mortal	MP:0001262:decr l	Psychosis proneness (hy	KEGG:Renin-angioter	Hypertension (ZScor
MP:0010768:mortality/aging		Sum eosinophil basophil	KEGG:TGF-beta signa	Cerebrovascular dis
		Heel bone mineral dens	PathC: pid:Nongenot	Polycystic ovary syn
MP:0005385:c-v sys_phen|MP:0005386:behavior/neurological			RxT:Neuronal Systen	obsessive-compulsiv
MP:0005384:cel_phen|MP:0005369:m		Type 2 diabetes (http://ı	PathC: netpath:TGF_	Cancer (ZScore: 3.26
; Phenotype: Pancrea	MP:0000443:abnl s	Pancreatitis (http://ww	KEGG:Protein digesti	Pancreatitis (ZScore
MP:0005387:immune sys_phen|MP:00		Lung cancer (http://ww	RxT:Immune System	Arthritis (ZScore: 4.1
MP:0010768:mortality/aging|MP:0002		Post bronchodilator FEV	KEGG:Wnt signaling p	Cancer (ZScore: 4.28
MP:0005380:embry	MP:0005385:c-v sy	Plasma renin activity lev	KEGG:PI3K-Akt signal	Laron syndrome (ZS
MP:0005387:immune sys_phen|MP:00		Esophageal cancer (alcc	PathC: inoh:IGF1 sigr	Cancer (ZScore: 4.64
MP:0005384:cel_phen|MP:0005376:ho		Birth weight (http://ww	KEGG:Adipocytokine	Diabetes mellitus (Z
MP:0005375:adipos	MP:0011100:prew	Body mass index (http:/	KEGG:Gastric cancer	Diabetes mellitus (Z
MP:0005369:muscle phenotype|MP:0(		Hand grip strength (http	KEGG:Insulin resistar	Type 2 diabetes mel
MP:0005369:muscle	MP:0011100:prew	Pancreatic cancer (http:	KEGG:Type II diabete	Pancreatic agenesis
		Type 2 diabetes (http://www.ebi.ac.uk/efo/EFO_0001360):8e-13		
		Type 2 diabetes (http://www.ebi.ac.uk/efo/E		Type 2 diabetes mel
MP:0005378:growth/size/body region		phenotype|MP:0005382	KEGG:B cell receptor	Schneckenbecken d
MP:0005391:vision/eye phenotype|MP:0005381:digestive/alim			KEGG:Maturity onset	Diabetes mellitus (Z
MP:0005376:homec	MP:0002965:incr c	Reaction time (http://w	KEGG:Maturity onset	Diabetes mellitus (Z
MP:0005390:skeleton phenotype|MP:(		Adolescent idiopathic sc	KEGG:PI3K-Akt signal	Diabetes mellitus (Z
MP:0005367:renal/urinary sys_phen|M		Type 2 diabetes (http://ı	KEGG:Maturity onset	Diabetes mellitus (Z
			KEGG:Retrograde en	Echinococcosis (ZSc
MP:0003631:nervou	MP:0002792:abnl r	Type 2 diabetes (http://ı	KEGG:Maturity onset	Cancer (ZScore: 5.30
MP:0005378:growth/size/body region		Type 2 diabetes (http://ı	KEGG:Gastric acid se	Cardiomyopathy (ZS
MP:0005379:endocrine/exocrine gland		Systolic blood pressure	KEGG:PI3K-Akt signal	Donohue syndrome
MP:0005370:liver/b	MP:0002968:incr c	Glycated hemoglobin le	KEGG:Central carbon	Diabetes mellitus (Z
MP:0003631:nervou	MP:0003795:abnl l	Adiponectin levels (http	KEGG:Adipocytokine	Obesity (ZScore: 7.9
MP:0005378:growth	MP:0005559:incr c	Diastolic blood pressure	KEGG:PI3K-Akt signal	Cancer (ZScore: 4.92
		Adolescent idiopathic scoliosis (http://www.e		Congenital myasthe
MP:0005378:growth/size/body region		HDL cholesterol (http://	KEGG:Cholesterol me	Diabetes mellitus (Z
MP:0005375:adipos	MP:0000198:decr c	General cognitive ability	RxT:Metabolism of R	Diabetes mellitus (Z
MP:0010768:mortality/aging|MP:0005		High density lipoprotein	KEGG:MAPK signalin	Cervical cancer (ZSc
MP:0005369:muscle phenotype|MP:0005382:craniofacial phen			KEGG:Pyrimidine me	Arterial calcification

osis (ZScore: 8.314)|Glycogen storage disease IXc (ZScore: 3.991)|Umbilical hernia (ZScore: 3.838)
ore: 3.570)|Biotinidase deficiency (ZScore: 3.445)|Greig cephalopolysyndactyly syndrome (ZScore: 3.071
disease (ZScore: 5.271)|Autosomal dominant nonsyndromic deafness 69 (ZScore: 3.935)|Congenital here
sparagine N-linked glycosylation|RxT:COPII-mediated vesicle transport|RxT:Cargo concentration in the E
sorder (ZScore: 7.593)|Atherosclerosis (ZScore: 7.315)|Coronary artery disease (ZScore: 7.239)|Diabete
(ZScore: 5.126)
e: 8.882)|C1 inhibitor deficiency (ZScore: 7.460)|Urticaria (ZScore: 5.396)|Lupus erythematosus (ZScore
r degeneration (ZScore: 5.126)|Membranoproliferative glomerulonephritis (ZScore: 3.562)|Basal lamina
e: 4.176)|Ostertagiasis (ZScore: 4.064)|Hepatitis C (ZScore: 4.036)|Cancer (ZScore: 3.888)|Hepatitis B (Z
e: 6.258)|Cutis laxa (ZScore: 5.011)|Lattice corneal dystrophy (ZScore: 4.674)|Cancer (ZScore: 4.035)|Ne
68)|Lung disease (ZScore: 5.038)|Stomach disease (ZScore: 4.702)|Liver disease (ZScore: 4.638)|Breast
ıe (ZScore: 5.840)|Cancer (ZScore: 3.962)|Atopic dermatitis (ZScore: 3.621)|Rosacea (ZScore: 3.336)|Ich
Score: 3.346)|Fraser syndrome (ZScore: 3.052)
.607)|Aromatic L-amino acid decarboxylase deficiency (ZScore: 4.877)|Parkinson's disease (ZScore: 4.48
460)|Common cold (ZScore: 4.553)|Cancer (ZScore: 4.222)|Hypereosinophilic syndrome (ZScore: 4.151)|
L30)|Eating disorder (ZScore: 5.306)|Diabetes mellitus (ZScore: 4.955)|Constipation (ZScore: 4.419)|Dia
re: 9.072)|Kidney disease (ZScore: 7.900)|Conn's syndrome (ZScore: 7.580)|Hypokalemia (ZScore: 6.964
ease (ZScore: 3.589)|Hemochromatosis (ZScore: 3.587)
ıdrome (ZScore: 7.551)|Hyperandrogenism (ZScore: 7.018)|Hypogonadism (ZScore: 6.297)|Obesity (ZSc
ve disorder (ZScore: 4.954)|Supratentorial primitive neuroectodermal tumor (ZScore: 3.765)|Trichotillor
66)
e: 7.272)|Pancreatic cancer (ZScore: 4.939)|Cystic fibrosis (ZScore: 4.505)|Autoimmune pancreatitis (ZSc
114)|Ulceroglandular tularemia (ZScore: 3.765)
80)|Gastrointestinal system disease (ZScore: 3.481)|Lung disease (ZScore: 3.010)
core: 6.927)|Acromegaly (ZScore: 5.368)|Turner syndrome (ZScore: 4.397)|Hypopituitarism (ZScore: 3.9
47)|Prostate disease (ZScore: 3.672)|Diabetes mellitus (ZScore: 3.589)|Angelman syndrome (ZScore: 3.3
Score: 6.672)|Hyperglycemia (ZScore: 6.027)|Obesity (ZScore: 6.010)|Hyperinsulinism (ZScore: 5.221)|F
Score: 6.045)|Obesity (ZScore: 3.884)|Hyperglycemia (ZScore: 3.177)|Cancer (ZScore: 3.055)
llitus (ZScore: 3.604)
(ZScore: 5.395)|Diabetes mellitus (ZScore: 5.300)|Hyperglycemia (ZScore: 4.049)|Insulinoma (ZScore: 3
llitus (ZScore: 3.161)
lysplasia (ZScore: 4.255)|Oculocerebrorenal syndrome (ZScore: 3.627)|Diabetes mellitus (ZScore: 3.461)
Score: 4.874)|Hyperglycemia (ZScore: 3.609)
Score: 4.927)|Hyperglycemia (ZScore: 3.106)
Score: 6.378)|Obesity (ZScore: 5.910)|Hyperglycemia (ZScore: 5.456)|Hyperinsulinism (ZScore: 5.246)|F
Score: 4.797)|Kidney disease (ZScore: 4.043)|Ovarian disease (ZScore: 4.023)|CAKUT (ZScore: 4.012)|Pr
ore: 6.391)|Leber hereditary optic neuropathy (ZScore: 6.133)|Fascioliasis (ZScore: 4.365)|Coenurosis (Z
04)|Diabetes mellitus (ZScore: 5.096)|Intestinal disease (ZScore: 4.489)|Liver disease (ZScore: 4.349)|St
Score: 7.069)|Short QT syndrome (ZScore: 6.087)|Beckwith-Wiedemann syndrome (ZScore: 5.205)|Atria
(ZScore: 5.730)|Diabetes mellitus (ZScore: 4.487)|Polycystic ovary syndrome (ZScore: 4.444)|Acanthosi
Score: 6.204)|Hyperglycemia (ZScore: 5.623)|Hyperinsulinism (ZScore: 4.669)|Hypoglycemia (ZScore: 4.
958)|Diabetes mellitus (ZScore: 7.289)|Atherosclerosis (ZScore: 6.012)|Nonalcoholic fatty liver disease (Z
20)|Diabetes mellitus (ZScore: 3.797)|Breast disease (ZScore: 3.515)|Ovarian disease (ZScore: 3.488)|Ob
enic syndrome 5 (ZScore: 3.239)|Ectopic pregnancy (ZScore: 3.005)
Score: 4.514)|Obesity (ZScore: 3.617)|Polycystic ovary syndrome (ZScore: 3.370)|Breast angiosarcoma
Score: 5.200)|Podoconiosis (ZScore: 3.199)|Obesity (ZScore: 3.014)
core: 7.596)|Carcinoma (ZScore: 5.027)|Vaginal cancer (ZScore: 3.413)|Adenocarcinoma in situ (ZScore:
ı of infancy (ZScore: 7.385)|Pseudoxanthoma elasticum (ZScore: 5.074)|Ankylosis (ZScore: 4.858)|Ricket

1)
editary endothelial dystrophy of cornea (ZScore: 3.899)|Leber congenital amaurosis (ZScore: 3.611)|Gla
ER|RxT:ER to Golgi Anterograde Transport|RxT:Immune System|RxT:Innate Immune System|RxT:Lysoso
s mellitus (ZScore: 6.336)|Obesity (ZScore: 5.748)

: 3.729)|Vascular disease (ZScore: 3.699)
ır drusen (ZScore: 3.299)|Complement deficiency (ZScore: 3.187)
ZScore: 3.646)
europathy (ZScore: 3.644)
disease (ZScore: 3.869)
thyosis (ZScore: 3.045)

9)|Meckel's diverticulum (ZScore: 4.428)|Occult macular dystrophy (ZScore: 4.199)
|Atopic dermatitis (ZScore: 4.030)
rrhea (ZScore: 3.903)
4)|Primary hyperaldosteronism (ZScore: 6.744)

ore: 5.726)|Anovulation (ZScore: 5.410)
mania (ZScore: 3.341)

core: 3.920)|Duodenal obstruction (ZScore: 3.638)

931)|Obesity (ZScore: 3.579)
313)|Breast disease (ZScore: 3.259)
Polycystic ovary syndrome (ZScore: 3.592)

3.697)|Fibrosarcoma (ZScore: 3.422)

Polycystic ovary syndrome (ZScore: 4.386)
rune belly syndrome (ZScore: 3.782)
ZScore: 4.200)|Leigh disease (ZScore: 4.145)
omach disease (ZScore: 3.600)
al fibrillation (ZScore: 5.081)|Cardiac arrest (ZScore: 4.536)
is nigricans (ZScore: 4.389)|Hyperinsulinism (ZScore: 4.239)
.598)|Hyperinsulinemic hypoglycemia (ZScore: 4.292)
ZScore: 5.917)|Lipid metabolism disorder (ZScore: 5.775)
besity (ZScore: 3.292)

(ZScore: 3.267)|Lipid metabolism disorder (ZScore: 3.084)

3.280)|Hemometra (ZScore: 3.255)
ts (ZScore: 4.698)|Hypophosphatasia (ZScore: 4.353)

ucoma (ZScore: 3.499)
ome Vesicle Biogenesis|RxT:Membrane Trafficking|RxT:Metabolism of Angiotensinogen to Angiotensins

|RxT:Metabolism of proteins|RxT:Neutrophil degranulation|RxT:Peptide hormone metabolism|RxT:Pos

t-translational protein modification|RxT:Transport to the Golgi and subsequent modification|RxT:Vesicl

e-mediated transport|RxT:trans-Golgi Network Vesicle Budding