Aus dem: Helmholtz Zentrum München und der Ludwig-Maximilians-Universität zu München

Dissertation

zum Erwerb des Doctor of Philosophy (Ph.D.)

an der Medizinischen Fakultät

der Ludwig-Maximilians-Universität zu München

Molecular Epidemiology of Cardiometabolic Risk Factors

vorgelegt von
Mohamed A. Elhadad

Aus
Ägypten

Jahr
2022

Mit Genehmigung der Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

First supervisor:	Prof. Dr. Annette Peters
Second supervisor:	Dr. Melanie Waldenberger
Third supervisor:	Prof. Dr. med. Wolfgang Koenig
Fourth supervisor:	PD. Dr. med. Moritz Sinner

Dean:	Prof. Dr. med. Thomas Gudermann

Datum der Verteidigung:	November 9, 2022

Affidavit

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN | Promotionsbüro Medizinische Fakultät | MMRS

Affidavit

Elhadad, Mohamed Ahmed

Surname, first name

I hereby declare that the submitted thesis entitled:

Molecular Epidemiology of Cardiometabolic Risk Factors

is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the dissertation presented here has not been submitted in the same or similar form to any other institution for the purpose of obtaining an academic degree.

Greifswald, 12.01.2023

place, date

Mohamed A. Elhadad

Signature doctoral candidate

Confirmation of congruency

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN | Promotionsbüro Medizinische Fakultät | MMRS

Confirmation of congruency between printed and electronic version of the doctoral thesis

Elhadad, Mohamed Ahmed

Surname, first name

I hereby declare that the submitted thesis entitled:

Molecular Epidemiology of Cardiometabolic Risk Factors

is congruent with the printed version both in content and format.

Greifswald, 12.01.2023

place, date

Mohamed A. Elhadad

Signature doctoral candidate

Table of content

List of abbreviations

CVD	cardiovascular disease
MetS	metabolic syndrome
HDL	high-density lipoprotein
LDL	low-density lipoprotein
CAD	coronary artery disease
PCSK9	Proprotein convertase subtilisin/kexin type 9
T2D	type 2 diabetes
MR	Mendelian randomization
KORA	Cooperative Health Research in the Augsburg Region
HUNT	The Trøndelag Health Study
DNA	Deoxyribonucleic acid
OGTT	oral glucose tolerance test
FDR	false discovery rate
LASSO	least absolute shrinkage and selection operator
AUC	Area Under the receiver operating characteristic curve
IV	instrumental variable
ESC	European society of cardiology
AHA	American Heart Association
PWAS	proteome-wide association study
BMI	body mass index
SHBG	sex hormone binding globulin
PYY	Peptide YY
TNFAIP6	Tumor necrosis factor-inducible gene 6 protein
CDNF	Cerebral dopamine neurotrophic factor
WIF1	WNT Inhibitory Factor 1
TGFbR3	Transforming Growth Factor Beta Receptor 3
GDRS	German diabetes risk score
sRAGE	Soluble advanced glycosylation end product-specific receptor

List of publications

1.1 Publications included in the thesis:

1. **Metabolic Syndrome and the Plasma Proteome: from Association to Causation; *Elhadad MA***, Wilson R, Zaghlool SB, Huth C, Gieger C, Grallert H, Graumann J, Rathmann W, Koenig W, Sinner MF, Hveem K, Suhre K, Thorand B, Jonasson C, Waldenberger M, Peters A; Cardiovasc Diabetol. 2021 May 20;20(1):111. doi: 10.1186/s12933-021-01299-2. PMID: 34016094; PMCID: PMC8138979.

2. **Deciphering the Plasma Proteome of Type 2 Diabetes; *Elhadad MA***, Jonasson C, Huth C, Wilson R, Gieger C, Matias P, Grallert H, Graumann J, Gailus-Durner V, Rathmann W, von Toerne C, Hauck SM, Koenig W, Sinner MF, Oprea TI, Suhre K, Thorand B, Hveem K, Peters A, Waldenberger M; Diabetes. 2020 Dec;69(12):2766-2778. doi: 10.2337/db20-0296. Epub 2020 Sep 14. PubMed PMID: 32928870.

1.2 Other publications:

1. **Integrative analysis of clinical and epigenetic biomarkers of mortality**; Huan T, Nguyen S, Colicino E, Ochoa-Rosales C, Hill WD, Brody JA, Soerensen M, Zhang Y, Baldassari A, ***Elhadad MA***, Toshiko T, Zheng Y, Domingo-Relloso A, Lee DH, Ma J, Yao C, Liu C, Hwang SJ, Joehanes R, Fornage M, Bressler J, van Meurs JBJ, Debrabant B, Mengel-From J, Hjelmborg J, Christensen K, Vokonas P, Schwartz J, Gahrib SA, Sotoodehnia N, Sitlani CM, Kunze S, Gieger C, Peters A, Waldenberger M, Deary IJ, Ferrucci L, Qu Y, Greenland P, Lloyd-Jones DM, Hou L, Bandinelli S, Voortman T, Hermann B, Baccarelli A, Whitsel E, Pankow JS, Levy D; Aging Cell. 2022 May 12:e13608. doi: 10.1111/acel.13608. Epub ahead of print. PMID: 35546478.

2. **Plasma Proteomics of Renal Function: A Trans-ethnic Meta-analysis and Mendelian Randomization Study**; Matías-García P, Wilson R, Guo Q, Zaghlool S, Eales J, Xu X, Charchar F, Dormer J, Maalmi H, Schlosser P, ***Elhadad M***, Nano J, Sharma S, Peters A, Fornoni A, Mook-Kanamori D, Winkelmann J, Danesh J, Di Angelantonio E, Ouwehand W, Watkins N, Roberts D, Petrera A, Graumann J, Koenig W, Hveem K, Jonasson C, Köttgen A, Butterworth A, Prunotto M, Hauck S, Herder C, Suhre K, Gieger C, Tomaszewski M, Teumer A, Waldenberger M; J Am Soc Nephrol. 2021 Jun 16;. doi: 10.1681/ASN.2020071070. [Epub ahead of print] PubMed PMID: 34135082; PubMed Central PMCID: PMC8425654.

3. **Differences in Biomarkers of Inflammation Between Novel Subgroups of Recent-Onset Diabetes**; Herder C, Maalmi H, Strassburger K, Zaharia OP, Ratter JM, Karusheva Y, ***Elhadad MA***, Bódis K, Bongaerts BWC, Rathmann W, Trenkamp S, Waldenberger M, Burkart V, Szendroedi J, Roden M; GDS

Group; Diabetes. 2021 May;70(5):1198-1208. doi: 10.2337/db20-1054. Epub 2021 Feb 19. PMID: 33608423.

4. **Association of Circulating Monocyte Chemoattractant Protein-1 Levels With Cardiovascular Mortality: A Meta-analysis of Population-Based Studies**; Georgakis MK, de Lemos JA, Ayers C, Wang B, Björkbacka H, Pana TA, Thorand B, Sun C, Fani L, Malik R, Dupuis J, Engström G, Orho-Melander M, Melander O, Boekholdt SM, Zierer A, ***Elhadad MA***, Koenig W, Herder C, Hoogeveen RC, Kavousi M, Ballantyne CM, Peters A, Myint PK, Nilsson J, Benjamin EJ, Dichgans M; JAMA Cardiol. 2021 May 1;6(5):587-592. doi: 10.1001/jamacardio.2020.5392. PMID: 33146689; PMCID: PMC8111478.

5. **Revealing the role of the human blood plasma proteome in obesity using genetic drivers**; Zaghlool SB, Sharma S, Molnar M, Matías-García PR, ***Elhadad MA***, Waldenberger M, Peters A, Rathmann W, Graumann J, Gieger C, Grallert H, Suhre K; Nat Commun. 2021 Feb 24;12(1):1279. doi: 10.1038/s41467-021-21542-4. PMID: 33627659; PMCID: PMC7904950.

6. **Whole Blood DNA Methylation Signatures of Diet Are Associated With Cardiovascular Disease Risk Factors and All-Cause Mortality;** Ma J, Rebholz CM, Braun KVE, Reynolds LM, Aslibekyan S, Xia R, Biligowda NG, Huan T, Liu C, Mendelson MM, Joehanes R, Hu EA, Vitolins MZ, Wood AC, Lohman K, Ochoa-Rosales C, van Meurs J, Uitterlinden A, Liu Y, ***Elhadad MA***, Heier M, Waldenberger M, Peters A, Colicino E, Whitsel EA, Baldassari A, Gharib SA, Sotoodehnia N, Brody JA, Sitlani CM, Tanaka T, Hill WD, Corley J, Deary IJ, Zhang Y, Schöttker B, Brenner H, Walker ME, Ye S, Nguyen S, Pankow J, Demerath EW, Zheng Y, Hou L, Liang L, Lichtenstein AH, Hu FB, Fornage M, Voortman T, Levy D; Circ Genom Precis Med. 2020 Aug;13(4):e002766. doi: 10.1161/CIRCGEN. 119.002766. Epub 2020 Jun 11. PubMed PMID: 32525743; PubMed Central PMCID: PMC7442697.

7. **Metabolites, Nutrients, and Lifestyle Factors in Relation to Coffee Consumption: An Environment-Wide Association Study; *Elhadad MA***, Karavasiloglou N, Wulaningsih W, Tsilidis KK, Tzoulaki I, Patel CJ, Rohrmann S; Nutrients. 2020 May 19;12(5). doi: 10.3390/nu12051470. PubMed PMID: 32438643; PubMed Central PMCID: PMC7284649.

8. **Epigenetics meets proteomics in an epigenome-wide association study with circulating blood plasma protein traits;** Zaghlool SB, Kühnel B, ***Elhadad MA***, Kader S, Halama A, Thareja G, Engelke R, Sarwath H, Al-Dous EK, Mohamoud YA, Meitinger T, Wilson R, Strauch K, Peters A, Mook-Kanamori DO, Graumann J, Malek JA, Gieger C, Waldenberger M, Suhre K; Nat Commun. 2020 Jan 3;11(1):15. doi: 10.1038/s41467-019-13831-w. PubMed PMID: 31900413; PubMed Central PMCID: PMC6941977.

9. **Circulating Monocyte Chemoattractant Protein-1 and Risk of Stroke: Meta-Analysis of Population-Based Studies Involving 17 180 Individuals**; Georgakis MK, Malik R, Björkbacka H, Pana TA, Demissie S, Ayers C, ***Elhadad MA***, Fornage M, Beiser AS, Benjamin EJ, Boekholdt SM, Engström G, Herder C, Hoogeveen RC, Koenig W, Melander O, Orho-Melander M, Schiopu

A, Söderholm M, Wareham N, Ballantyne CM, Peters A, Seshadri S, Myint PK, Nilsson J, de Lemos JA, Dichgans M; Circ Res. 2019 Sep 27;125(8):773-782. doi: 10.1161/CIRCRESAHA.119.315380. Epub 2019 Sep 3. Erratum in: Circ Res. 2020 Jan 3;126(1):e9. PMID: 31476962; PMCID: PMC6763364.

10. **Abstract 21: Deciphering the Plasma Proteome of Type 2 Diabetes; *Elhadad MA***, Jonasson C, Huth C, Wilson R, Thorand B, Grallert H, Graumann J, Rathmann W, Koenig W, Sinner MF, Gieger C, Suhre K, Hveem K, Peters A, Waldenberger M; Circulation. 2020; 141:A21; doi: 10.1161/circ.141.suppl_1.21.

11. **P4769, Using the plasma proteome to decipher metabolic syndrome pathophysiology and discover a diagnostic biomarker panel; *Elhadad MA***, Wilson R, Zaghlool S, Huth C, Kriebel J, Grallert H, Rathmann W, Graumann J, Suhre K, Peters A, Gieger C, Waldenberger M, German Research Center for Cardiovascular Disease (DZHK), Partner site Munich Heart Alliance, Germany; European Heart Journal, Volume 39, Issue suppl_1, August 2018, ehy563.P4769, doi: 10.1093/eurheartj/ehy563.P4769.

12. **Research Collaborations: A guide for early career researchers by early career researchers; *Elhadad MA*** et al.; University of Edinburgh; July 2019; doi: 10.13140/RG.2.2.36095.36004

Author’s contribution to the publications

1.3 Contribution to paper I

The study “Deciphering the Plasma Proteome of Type 2 Diabetes” was published in Diabetes in 2020. The research was designed by Mohamed Elhadad in consultation with Prof. Annette Peters and Dr. Melanie Waldenberger. Mohamed Elhadad performed the statistical analysis, interpreted the results, wrote the manuscript, and accompanied the publication process as corresponding author. All co-authors critically revised and approved the manuscript.

1.4 Contribution to paper II

The study “Metabolic Syndrome and the Plasma Proteome: From Association to Causation” was published in Cardiovascular Diabetology in 2021. The research was designed by Mohamed Elhadad in consultation with Prof. Annette Peters and Dr. Melanie Waldenberger. Mohamed Elhadad performed the statistical analysis, interpreted the results, wrote the manuscript, and accompanied the publication process as corresponding author. All co-authors critically revised and approved the manuscript.

2. Introductory summary

2.1 Background

Atherosclerosis, in particular ischemic heart disease and stroke, accounted for 27 % of the mortality worldwide in 2019 making cardiovascular disease (CVD) the most common cause of morbidity and mortality worldwide (1). These numbers have been increasing in recent decades driven by the higher life expectancies, which we were able to achieve though the advances in medical knowledge and technology, and the prevailing sedentary lifestyle especially in developed countries (2). Similar, older age combined with sedentary lifestyle increased the prevalence of cardio-metabolic risk factors like obesity and diabetes (2).

In the early stages of cardiovascular research emphasis was put on the discovery of the mechanisms governing the functions of the heart and its associated circulatory systems like the discovery of the pulmonary circulation by Ibn al Nafis in the 13th century (3). The twentieth century saw the introduction of cohort studies (4), which have since been extensively applied in cardiovascular research to uncover its risk factors and the mechanisms involved in disease development and progression (5).

Early cohort studies of CVD like the Framingham study unveiled the association of important risk factors with CVD like hypertension (6), and blood cholesterol levels (6, 7). These findings have been repeatedly replicated and expanded upon in other studies including reports on the role of cardiometabolic risk factors in CVD's pathogenesis like obesity (8), diabetes (8), and metabolic syndrome (9).

Diabetes is an established risk factor of CVD (10). Diabetic patients do not only have higher risk of developing CVD among other diabetic complications, but they are also more prone to much severer forms of the disease (10). These observations are the background behind the current clinical practice guidelines emphasizing on tight control of diabetes in CVD high risk patients (11, 12).

Further studies on the health effects of diabetes draw the attention to a peculiar clustering of metabolic risk factors, leading to a much greater risk of CVD compared to diabetes alone (13). This clustering was called many names, for example syndrome x (13) and later the metabolic syndrome (MetS) (9). The definition of the syndrome changed over time and it is now defined as having at least three of five possible components namely abdominal obesity, high blood pressure, decreased levels of high-density lipoprotein (HDL), increased levels of low-density lipoprotein (LDL) and higher fasting glucose levels (9, 14).

Studies on the role of the different metabolic risk factors of CVD have suggested insulin resistance and its subsequent dysregulation in carbohydrate and lipid metabolism as a common pathway. However, the constellation of risk factors like that observed in MetS remains elusive to explanation, pushing more of the research efforts toward the study of single risk factors of CVD like LDL (11).

Discovery of the correlation between cholesterol level and atherosclerosis like coronary artery disease (CAD), opened the door to studies into the lipid subtypes and their roles in CVD (15). Driven by the reported correlation between cholesterol level and atherosclerosis like CAD, research on the role of LDL in the pathophysiology of atherosclerosis culminated with the identification of new drugs and preventive measures, the most successful of which so far is Proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitors (16, 17).

The introduction of molecular epidemiology in CVD research had a great impact on the field and our current understanding of CVD pathophysiology. Different layers of the molecular cascade offer different insights into cardiovascular physiology and disease. Proteins, the products of genes, are the workhorse inside the cell. Modulated expression of genes results in differences in protein concentration with subsequent modulations of cellular functions (18). Proteomics reflect not only gene expression, translation and post-translational modifications but also sheds the light on protein-protein interactions and the interaction with other molecular layers of the physiological and pathological processes (18-20).

Proteomics has been under constant development with recent advances in mass spectrometry allowing for the measurement of much more proteins in one sample (19). The introduction of aptamer-based proteomics allows us to measure even more proteins in one small sample (21), with some platforms measuring more than 5000 proteins in each sample (20).

In the current thesis, we aimed to use molecular level proteomics data investigating the cardiometabolic risk factors diabetes and MetS. Benefitting from the recent advances in proteomics, we used the aptamer-based protein measurements to unveil new associations with our outcomes. Additionally, we integrate available genetic association studies' results with our results to run a two-sample Mendelian randomization analyses aiming to prioritize our results and/or potentially uncover causal relations.

2.2 Aims of this study

In this study, we aimed to investigate the associations between plasma measured proteins and cardio-metabolic disorders namely type 2 diabetes (T2D) and MetS

cross-sectionally and - when follow up data was available - longitudinally. We then assessed the replicated results regarding pathway classification. Starting with the proteome-wide significant proteins, we intended to build protein risk scores to predict incident disease as well as assess their use as individual biomarkers. Furthermore, we utilized the two sample Mendelian randomization (MR) technique to overcome the limitation of observational studies in determining the direction of the association and in differentiating correlation from causation. Our MR analyses results could also be used as a prioritization tool. Additionally, we applied the MR analysis in both causality directions, when applicable, to conclude effect directions (22, 23).

2.3 Methods

2.3.1 Participating studies

For the first author papers, we used the data of two main studies in our analyses: the KORA (Cooperative Health Research in the Augsburg Region) cohort and the HUNT study (The Trøndelag Health Study). Detailed description of both studies has been reported before (24, 25). In short, the KORA is a cohort study with participants recruited from southern Germany, while HUNT included participants recruited from the county of Trøndelag in Norway (22, 23).

Both KORA and HUNT were reviewed and approved by respective local ethics committees. Both studies adhered to the rules of Helsinki with written informed consent collected from each participant. In both studies, each participant was extensively interviewed and examined to collect clinical and demographic data. Blood was also collected and stored for later OMICs measurement (22, 23).

As a discovery study, we used the subsample of the KORA F4 survey with available proteomics data for the cross-sectional analysis and the follow up survey KORA FF4 for the longitudinal analysis. As a replication study, we used the HUNT 3 study for the cross-sectional analysis and extracted follow up data from hospital and primary care records for the follow up analyses (22, 23).

Detailed information on the sample size and selection is available in the methods section of each paper (22, 23).

2.3.2 Proteomics data

Proteomics data used in both cohorts were measured using the aptamer-based technique “SOMAscan” (21). Technical information on the platform has been published elsewhere (21). In short, the aptamer-based technique works by having one aptamer that specifically binds to one protein (21). Multiple washing steps are then applied to

increase specificity by washing non-specific protein-aptamer complexes (21). The aptamers are then quantified as a proxy of protein concentration on a Deoxyribonucleic acid (DNA) microarray in relative fluorescence units (21).

To standardize the proteomics measurement, we log2 transformed the data and then standardized it by subtracting the mean and dividing by the standard deviation (22, 23).

2.3.3 Outcome definition, confounders, and association analysis

In both manuscripts, we aimed to utilize definitions that reflect clinically applied ones. Extensive details on outcome definition and choice of confounders could be found in the methods section of each manuscript (22, 23). Briefly, for type 2 diabetes, depending on availability, we used the validated clinically defined variables. In KORA, the self-reported type 2 diabetes variable was validated using clinical available data as well as oral glucose tolerance test (OGTT) measurements (26). In HUNT, self-reported variable, due to lack of OGTT measurements, was validated using primary health and hospital data (23).

For MetS we used the definition as per Alberti et al. (14). Participants were categorized as MetS if they met 3 out of 5 predefined criteria. The criteria were having 1) waist circumference of at least 94 cm in men and 80 cm in women, 2) fasting triglyceride levels of at least 150 mg/dl, 3) HDL lower than 40 in men and 50 in women, 4) blood pressure of at least 130mmHg systolic or 85 diastolic or previously diagnosed hypertension and 5) fasting glucose level of at least 100 mg/dl or previously diagnosed diabetes (22).

To calculate the association between proteins and respective outcome of each manuscript, we applied logistic regression models (22, 23).For each outcome we ran a proteome-wide analysis adjusted for a predefined set of confounding factors with one model per each protein (22, 23).

In manuscript one, we applied the false discovery rate (FDR) (27) and in the second manuscript we applied Bonferroni correction to account for multiple testing.

2.3.4 Biomarker discovery analyses

We built and tested the performance of protein risk scores as predictive tools for each of the studied outcomes (22, 23). First, we used the least absolute shrinkage and selection operator (LASSO) (28) to statistically select proteins with the highest predictive value. We then later tested the performance of our newly developed scores using the Area Under the receiver operating characteristic curve (AUC) statistic (29).

2.3.5 Integration of molecular data to infer causality

Finally, we applied two-sample Mendelian randomization to explore causality of our results. Details of this analytical methods have been extensively described elsewhere (30, 31).

In short, Mendelian randomization is an instrumental variable analysis, which was adopted from economics to biology (32). Genotype, which is randomly allocated at birth, is used as an instrumental variable (IV) in an analysis mimicking clinical trials to infer causality of an exposure of interest on an outcome of interest (31, 32). The IV, which is significantly associated with the exposure is used to estimate the causal effect of that exposure on the outcome (30-32).

Multiple assumptions need to be met in MR as illustrated in Figure 1: the IV must be associated with the exposure. If the IV is associated with the outcome, it must be only associated through the exposure and not through any other pathways or confounders (30, 31, 33).

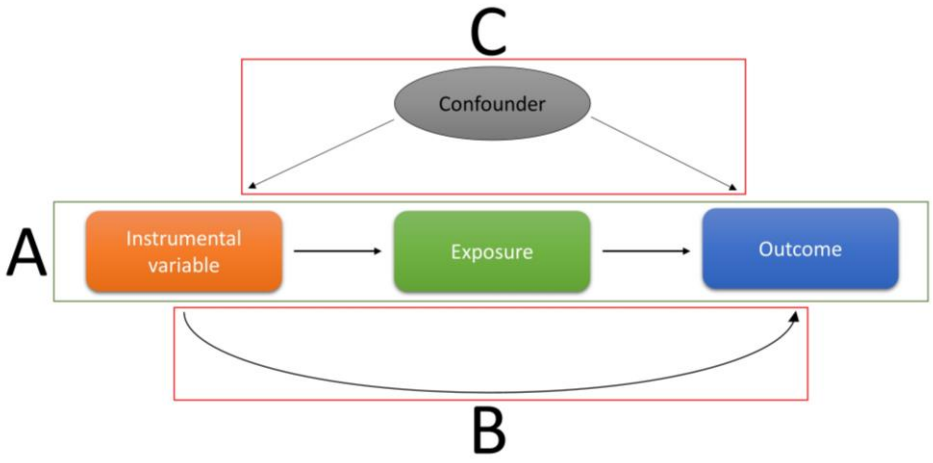

Figure 1: assumptions of Mendelian randomization analysis. A) The instrumental variable is associated with the exposure, associates with the outcome only through the exposure. B) The instrumental variable does not associate with the outcome through other independent pathways. C) The instrumental variable does not associate with the outcome through confounders.

2.4 Summary to article 1: Proteomics of type 2 diabetes

Diabetes is one of the well-known early-identified risk factors of CVD (34). Diabetic patients don't only have a higher prevalence of CVD than the normal population, they are also susceptible to the development of much severer forms of CVD (34). Therefore, a great emphasis on the tight control of diabetes has been recommended by CVD guidelines like those from the European society of cardiology (ESC) (35) and the American Heart Association (AHA) (11).

Through the application of proteomics, we aimed to unravel the proteins associated with both prevalent and incident T2D, explore their utility as biomarkers and to investigate their causal framework (23).

For proteome-wide association study (PWAS), we used logistic regression with T2D as the outcome in both cross-sectional and longitudinal analyses (23). We applied one model per protein and all models were adjusted for age, sex, body mass index (BMI), smoking status, and current hypertension at baseline (23).

For prevalent T2D, 24 out of the 85 KORA FDR significant proteins replicated in HUNT (23). With incident T2D, 3 out of the 10 KORA FDR significant proteins replicated in HUNT (23). Replicated proteins included proteins previously reported to be associated with T2D and successfully identified new proteins (23).

Aminoacylase-1, the only protein overlapping replicated results of prevalent and incident T2D, is one of the newly identified proteins (23). Aminoacylase-1, an enzyme that deacylates N-acylated L-amino acids other than proline and aspartate (36), has been found to be increased in hepatocytes (37) and decreased in omental lipocytes (38) of obese individuals, pointing to its potential and complex role in obesity. Moreover, aminoacylase-1 was found to be associated with the Framingham risk score, which predicts cardiovascular risk (39). These results point to the role of aminoacylase-1 as a potential link between both entities. Further investigation of the role of aminoacylase-1 in obesity, T2D and CVD is required.

The replicated proteins with incident T2D included previously reported associations namely insulin growth factor binding protein 2 and growth hormone receptor (40, 41).

Likewise, prevalent T2D protein associations replicated known association including mass spectrometry results like sex hormone binding globulin (SHBG) (42), renin (43) and gelsolin (44).

Analysis of prevalent type 2 diabetes also uncovered novel associations including Peptide YY (PYY), Tumor necrosis factor-inducible gene 6 protein (TNFAIP6), Cerebral dopamine neurotrophic factor (CDNF), WNT Inhibitory Factor 1 (WIF1) and Transforming Growth Factor Beta Receptor 3 (TGFbR3). Some of these have been reported to be associated with diabetic complications like CVD and atherosclerosis for example TNFAIP6, CDNF, WIF1, TGFbR3, and PYY (45-49).

To explore the predictive power of our newly discovered protein-associations, we modified the German diabetes risk score (GDRS) model (50) using variables available in our data. We then compared our protein extended model to the modified GDRS model. In general, the protein extended model showed only modest improvement over the original model, suggesting that protein measurement is not yet suitable for clinical application (23). While the cost of protein measurement using aptamer-based

techniques is not high, the added predictive value is not significantly high enough to justify such cost.

Utilizing Mendelian randomization analyses, we could infer causality of the relation between several proteins -with available instrumental variables- and T2D. SHBG showed a suggestive harmful causal effect on T2D (23). SHBG is reportedly implicated in the pathogenesis of T2D particularly insulin resistance (42). On the other direction of causality, we could show potential causal effects of T2D on Cathepsin Z and renin with an effect direction identical to that of observational results (23).

A strength of this study is the use of SOMAscan proteomics data, which enables the measurement of more than a thousand proteins representative of a wide range of pathways (21). We were able to replicate our results in the HUNT study allowing for better generalizability. We used readily available genetic association data to infer causality of our protein associations as well as the causal effect direction (23).

We tried to our best to address our study's limitations. The SOMAscan platform -as an aptamer-based platform- is liable to probe cross reactivity and non-specific binding (51). We confirmed that our replicated proteins have not been reported to have such problems (23). We could not replicate the rigorous T2D definition applied in KORA in the HUNT study due to the lack of oral glucose tolerance test data in HUNT. However, we used available data from hospital and primary care records to validate the T2D variable in HUNT (23).

Proteins are dynamic molecular players and are therefore under constant changes. Because of the nature of our study design, we were not able to investigate these dynamic changes. While we applied rigorous analytic strategies in our causal analyses, these results should be cautiously interpreted with the problems of MR analyses including but not limited to pleiotropy in mind (30, 52).

2.5 Summary to article 2: Proteomics of metabolic syndrome

The aim of this second manuscript was to investigate the link between proteomics of MetS, which comprises a group of metabolic risk factors of CVD (22). Insulin resistance, visceral obesity, a harmful lipid profile in the form of increased LDL and decreased HDL and increased blood pressure are the components of MetS. All of which are proved risk factors of CVD and in particular atherosclerosis, plaque formation and instability (53).

Although the syndrome has been described decades ago, its definition has been under constant change with different experts and societies proposing different defining components and cutoff points. To overcome this, we used the harmonized definition of MetS in both studies (14).

Our proteome-wide analyses with prevalent MetS yielded 116 protein associations, of which we could replicate 53 in the HUNT study. Proteome-wide study with incident MetS yielded 14 proteins. Due to data unavailability in the HUNT study, we were not able to replicate the results of incident MetS (22).

The associations of incident MetS overlapped to a great extent with those associated with prevalent MetS. Soluble advanced glycosylation end product-specific receptor (sRAGE) was exclusively associated with incident MetS (22).

Our results included known associations like leptin (54), thus replicating the previous reports of MetS protein associations. We also report new associations like NTR domain-containing protein 2 and endoplasmic reticulum protein 29 (22).

We evaluated the role of our associated proteins as potential biomarkers using LASSO to select a subset of proteins with the highest predictive value. Our selected diagnostic model comprised eight proteins and had an AUC of 0.75 in KORA (22).

Finally, we studied the causal framework governing our observational results by applying two-sample Mendelian randomization techniques. We revealed probable causal effects of apolipoprotein E2, apolipoprotein B and proto-oncogene tyrosine-protein kinase receptor on MetS (22).

2.6 Research potential

In the current thesis we demonstrated the potential of high throughput proteomics and its integration with genetic data. The same approach could be applied to larger datasets and proteomics assays increasing power and providing a wider pathway coverage.

Analyses of population subgroups of cardiometabolic risk factors, for example investigating the proteomic association and causal framework of people with CVD and diabetes compared to those without diabetes, would offer a much-needed insight into the pathophysiological mechanisms behind the observed differences in CVD-risk attributed to each of these subgroups.

Additionally, there is great potential in the integration of metabolomics data with the proteomics, which would complement information obtained from proteomics data analysis, although such integration could prove hard to interpret especially due to the limitations of currently available statistical methods (55).

The combination of different layers of the molecular data to use in individualized medicine to identify and correctly label patient subgroups and identify new treatments and preventive measures specifically tailored for these subgroups, might be of great value to maximize these treatments' effect and avoiding or reducing their side effects.

2.7 Conclusion

Though the use of state-of-the-art proteomics' measurement technique, we were able to investigate the association of the metabolic risk factors of CVD: type 2 diabetes and MetS. Our results replicated previously published association, thus reflecting the integrity of the aptamer-based proteomics technique and revealed new protein associations like Aminoacylase-1 with T2D and NTR domain-containing protein 2 with MetS.

The overlap between the T2D and MetS associated proteins underlines the potential involvement of common pathways in the pathophysiological processes leading to both diseases and their associated complications.

By integrating the genetic data with proteomics data, we were able to investigate the causality framework of our results and to prioritize them.

References

1. World Health Organization. Fact Sheet: The top 10 causes of death. 2020 [Available from: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.

2. Health Effects of Overweight and Obesity in 195 Countries over 25 Years. New England Journal of Medicine. 2017;377(1):13-27.

3. Meyerhof M. Ibn An-Nafîs (XIIIth Cent.) and His Theory of the Lesser Circulation. Isis. 1935;23(1):100-20.

4. Doll R. Cohort studies: history of the method. I. Prospective cohort studies. Soz Praventivmed. 2001;46(2):75-86.

5. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet (London, England). 2014;383(9921):999-1008.

6. Kannel WB, Dawber TR, Kagan A, Revotskie N, Stokes J, 3rd. Factors of risk in the development of coronary heart disease--six year follow-up experience. The Framingham Study. Annals of internal medicine. 1961;55:33-50.

7. Kannel WB, Dawber TR, Friedman GD, Glennon WE, McNamara PM. Risk Factors in Coronary Heart Disease. An Evaluation of Several Serum Lipids as Predictors of Coronary Heart Disease; The Framingham Study. Annals of internal medicine. 1964;61:888-99.

8. Piché M-E, Tchernof A, Després J-P. Obesity Phenotypes, Diabetes, and Cardiovascular Diseases. Circulation Research. 2020;126(11):1477-500.

9. Hanefeld M, Schaper F, Ceriello A. Geschichte und Definition(en) des metabolischen Syndroms. Der Internist. 2007;48(2):117-25.

10. Low Wang CC, Hess CN, Hiatt WR, Goldfine AB. Clinical Update: Cardiovascular Disease in Diabetes Mellitus: Atherosclerotic Cardiovascular Disease and Heart Failure in Type 2 Diabetes Mellitus - Mechanisms, Management, and Clinical Considerations. Circulation. 2016;133(24):2459-502.

11. Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation. 2019;140(11):e596-e646.

12. Visseren FLJ, Mach F, Smulders YM, Carballo D, Koskinas KC, Bäck M, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice: Developed by the Task Force for cardiovascular disease prevention in clinical practice with representatives of the European Society of Cardiology and 12 medical societies With the special contribution of the European Association of Preventive Cardiology (EAPC). European Heart Journal. 2021;42(34):3227-337.

13. Reaven GM. Role of Insulin Resistance in Human Disease. Diabetes. 1988;37(12):1595-607.

14. Alberti KG, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation;

International Atherosclerosis Society; and International Association for the Study of Obesity. Circulation. 2009;120(16):1640-5.

15. Ding M, Rexrode KM. A Review of Lipidomics of Cardiovascular Disease Highlights the Importance of Isolating Lipoproteins. Metabolites. 2020;10(4).

16. Abifadel M, Varret M, Rabes JP, Allard D, Ouguerram K, Devillers M, et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nature genetics. 2003;34(2):154-6.

17. Sabatine MS. PCSK9 inhibitors: clinical evidence and implementation. Nat Rev Cardiol. 2019;16(3):155-65.

18. Aslam B, Basit M, Nisar MA, Khurshid M, Rasool MH. Proteomics: Technologies and Their Applications. Journal of Chromatographic Science. 2017;55(2):182-96.

19. Nesvizhskii AI, Aebersold R. Interpretation of shotgun proteomic data: the protein inference problem. Molecular & cellular proteomics : MCP. 2005;4(10):1419-40.

20. Palstrøm NB, Matthiesen R, Rasmussen LM, Beck HC. Recent Developments in Clinical Plasma Proteomics-Applied to Cardiovascular Research. Biomedicines. 2022;10(1).

21. Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. PloS one. 2010;5(12):e15004.

22. Elhadad MA, Wilson R, Zaghlool SB, Huth C, Gieger C, Grallert H, et al. Metabolic syndrome and the plasma proteome: from association to causation. Cardiovascular diabetology. 2021;20(1):111.

23. Elhadad MA, Jonasson C, Huth C, Wilson R, Gieger C, Matias P, et al. Deciphering the Plasma Proteome of Type 2 Diabetes. Diabetes. 2020;69(12):2766-78.

24. Krokstad S, Langhammer A, Hveem K, Holmen TL, Midthjell K, Stene TR, et al. Cohort Profile: the HUNT Study, Norway. International journal of epidemiology. 2013;42(4):968-77.

25. Holle R, Happich M, Löwel H, Wichmann HE. KORA--a research platform for population based health research. Gesundheitswesen. 2005;67 Suppl 1:S19-25.

26. Huth C, von Toerne C, Schederecker F, de las Heras Gala T, Herder C, Kronenberg F, et al. 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(4):409-22.

27. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300.

28. Tibshirani R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society Series B (Methodological). 1996;58(1):267-88.

29. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837-45.

30. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37(7):658-65.

31. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife. 2018;7.

32. 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. 2016;45(6):1717-26.

33. Holmes MV, Ala-Korpela M, Smith GD. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. Nature Reviews Cardiology. 2017;14:577.

34. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, et al. Heart Disease and Stroke Statistics—2021 Update. Circulation. 2021;143(8):e254-e743.

35. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts)Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). Eur Heart J. 2016;37(29):2315-81.

36. Mitta M, Kato I, Tsunasawa S. The nucleotide sequence of human aminoacylase-1. Biochimica et biophysica acta. 1993;1174(2):201-3.

37. Caira S, Iannelli A, Sciarrillo R, Picariello G, Renzone G, Scaloni A, et al. 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(1):672-82.

38. Perez-Perez R, Garcia-Santos E, Ortega-Delgado FJ, Lopez JA, Camafeita E, Ricart W, et al. Attenuated metabolism is a hallmark of obesity as revealed by comparative proteomic analysis of human omental adipose tissue. Journal of proteomics. 2012;75(3):783-95.

39. Ngo D, Sinha S, Shen D, Kuhn EW, Keyes MJ, Shi X, et al. Aptamer-Based Proteomic Profiling Reveals Novel Candidate Biomarkers and Pathways in Cardiovascular Disease. Circulation. 2016;134(4):270-85.

40. Wittenbecher C, Ouni M, Kuxhaus O, Jähnert M, Gottmann P, Teichmann A, et al. Insulin-Like Growth Factor Binding Protein 2 (IGFBP-2) and the Risk of Developing Type 2 Diabetes. Diabetes. 2018:db180620.

41. Lu M, Flanagan JU, Langley RJ, Hay MP, Perry JK. Targeting growth hormone function: strategies and therapeutic applications. Signal Transduction and Targeted Therapy. 2019;4(1):3.

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(1):32-40.

43. 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(17):e009890.

44. Khatri N, Sagar A, Peddada N, Choudhary V, Chopra BS, Garg V, et al. 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.

45. 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(2):465.

46. Zhang G-L, Wang L-H, Liu X-Y, Zhang Y-X, Hu M-Y, Liu L, et al. 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(7):1905.

47. Ress C, Paulweber M, Goebel G, Willeit K, Rufinatscha K, Strobl A, et al. Circulating Wnt inhibitory factor 1 levels are associated with development of cardiovascular disease. Atherosclerosis. 2018;273:1-7.

48. Sun F, Li X, Duan W-Q, Tian W, Gao M, Yang J, et al. Transforming Growth Factor-β Receptor III is a Potential Regulator of Ischemia-Induced Cardiomyocyte Apoptosis. Journal of the American Heart Association. 2017;6(6):e005357.

49. 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(6):648-52.

50. Paprott R, Mühlenbruch K, Mensink GBM, Thiele S, Schulze MB, Scheidt-Nave C, et al. 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(1):e000280.

51. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature. 2018;558(7708):73-9.

52. Zheng J, Baird D, Borges MC, Bowden J, Hemani G, Haycock P, et al. Recent Developments in Mendelian Randomization Studies. Current epidemiology reports. 2017;4(4):330-45.

53. Libby P. The changing landscape of atherosclerosis. Nature. 2021;592(7855):524-33.

54. Lind L, Elmståhl S, Ingelsson E. Cardiometabolic Proteins Associated with Metabolic Syndrome. Metab Syndr Relat Disord. 2019;17(5):272-9.

55. Jendoubi T. Approaches to Integrating Metabolomics and Multi-Omics Data: A Primer. Metabolites. 2021;11(3).

3. Paper I:

Deciphering the Plasma Proteome of Type 2 Diabetes

Mohamed A. Elhadad; Christian Jonasson; Cornelia Huth; Rory Wilson; Christian Gieger; Pamela Matias; Harald Grallert; Johannes Graumann; Valerie Gailus-Durner; Wolfgang Rathmann; Christine von Toerne; Stefanie M. Hauck; Wolfgang Koenig; Moritz F. Sinner; Tudor I. Oprea; Karsten Suhre; Barbara Thorand; Kristian Hveem; Annette Peters; Melanie Waldenberger

Diabetes. 2020 Dec;69(12):2766-2778

https://doi.org/10.2337/db20-0296

2766 **Diabetes** Volume 69, December 2020

Deciphering the Plasma Proteome of Type 2 Diabetes

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

Diabetes 2020;69:2766–2778 | https://doi.org/10.2337/db20-0296

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 >1,000 plasma proteins in the Cooperative Health Research in the Region of Augsburg (KORA) F4 cohort (n = 993, 110 cases), with subsequent replication in the third wave of the Nord-Trøndelag Health Study (HUNT3) 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 bidirectional Mendelian randomization (MR) analysis to prioritize our results. Association analysis of prevalent type 2 diabetes revealed 24 replicated proteins, of which 8 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.

[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, Munich, Germany
[4]K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, Norwegian University of Science and Technology, Trondheim, Norway
[5]HUNT Research Center, Department of Public Health, Norwegian University of Science and Technology, Levanger, Norway
[6]German Center for Diabetes Research (DZD), München-Neuherberg, Germany
[7]Biomolecular Mass Spectrometry, Max Planck Institute for Heart and Lung Research, Bad Nauheim, 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, Munich, Germany
[13]Institute of Epidemiology and Medical Biometry, University of Ulm, Ulm, Germany
[14]Department of Medicine I, University Hospital Munich, Ludwig Maximilian University, Munich, Germany
[15]Department of Internal Medicine and UNM Comprehensive Cancer Center, University of New Mexico School of Medicine, Albuquerque, NM
[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]Department of Biophysics and Physiology, Weill Cornell Medicine - Qatar, Education City, Doha, Qatar
[19]Institute of Medical Information Sciences, Biometry and Epidemiology, Ludwig Maximilian University, Munich, Germany

Corresponding authors: Melanie Waldenberger, waldenberger@helmholtz-muenchen.de, and Mohamed A. Elhadad, mohamed.elhadad@helmholtz-muenchen.de

Received 2 April 2020 and accepted 28 August 2020

This article contains supplementary material online at https://doi.org/10.2337/figshare.12896426.

Type 2 diabetes is a significant cause of morbidity and mortality, with an estimated worldwide prevalence of 463 million patients, one-half of whom are undiagnosed (1). It is a complex, multifactorial disease characterized by an interplay of both genetic and nongenetic factors that lead 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 patients with diabetes with healthy control subjects have 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, β-carotene, N-terminal proB-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 1,095 plasma proteins in the Cooperative Health Research in the Region of Augsburg (KORA) study. We replicate our results in the independent Nord-Trøndelag Health Study (HUNT) study and 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 bidirectional MR analysis (11) to assess causality and prioritize the newly discovered relationships.

RESEARCH DESIGN AND METHODS

Study Populations

KORA Cohort

The KORA study comprises independent samples from Augsburg in southern Germany (12). In the current study, we used a subsample of 1,000 individuals randomly selected from the participants of the KORA F4 survey (N = 3,080, performed 2006–2008) with deep phenotyping data (n = 1,800) (13). Detailed clinical and sociodemographic 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 (Berlin, Germany) reviewed and approved the study, and all participants gave written informed consent.

HUNT Cohort

HUNT is a prospective population-based cohort from Nord-Trøndelag County in Norway (14). We used the HUNT3 survey (n = 1,117 with proteomics measurements, performed 2006–2008) for the validation of the KORA study results. The HUNT study collected detailed sociodemographic 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 nonfasting 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, were incubated. After washing steps, bead-bound proteins were biotinylated, and complexes comprising biotinylated target proteins and fluorescence-labeled aptamers were photocleaved off the bead support and pooled. Following recapture on streptavidin beads and further washing steps, aptamers were eluted and quantified as a proxy to protein concentration by hybridization to custom arrays of aptamer-complementary oligonucleotides. On the basis of standard samples included on each plate, the resulting raw intensities were processed using a data analysis workflow that included hybridization normalization, median signal normalization, and signal calibration to control for interplate 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 version 3.2), 29 failed SOMAscan quality control. We also removed the five probes recommended by the SOMAscan assay change log issued on 22 December 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 (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 ICD-10 code E11 and the International Classification of Primary Care, Second Edition, 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 1 h/week as physically active (more details are available in the Supplementary Material). 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 International Classification of Primary Care, Second Edition, 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 with them to their study center visit. Using database software (17), medications were identified using Anatomical Therapeutic Chemical 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 $\log_2$ transformation and (0, 1) standardization by subtracting the per-cohort mean and dividing by the per-cohort SD to allow easier interpretation of the odds ratios (ORs) 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., 1,095 models per outcome) and adjusted for the potential confounders age, sex, 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 on the basis of the number of significant proteins in KORA. To examine whether antidiabetic 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 through the Pharos (18) platform, experimental Gene Ontology (19) terms, Kyoto Encyclopedia of Genes and Genomes (KEGG) (20) pathways, human disease association data from the GWAS Catalog (21), Online Mendelian Inheritance in Man (22), and the text-mining DISEASES platform (23) as well as through phenotype data from the corresponding mouse ortholog 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 10 proteins, of which 1 failed quality control in HUNT. We used KORA as a training data set and HUNT as a test data set and used an adapted version of GDRS, a diabetes risk score that was trained in 21,845 participants of the European Prospective Investigation into Cancer and Nutrition Potsdam (EPIC-Potsdam) study with a mean follow-up time of 7 years, as a benchmark (9). More details on the GDRS are available in the Supplementary Material.

Adaptation of the GDRS was necessary because some of the variables were missing from one or both cohorts. We defined smoking status using only information on current and former smoking per se without regard to the number 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 child with diabetes. We calculated the risk of a positive family history by averaging the original GDRS of having one parent, both parents, or at least one sibling with type 2 diabetes ($[56 + 106 + 48] / 3 = 70$). Our final adapted GDRS was calculated as follows: $5.1 \times$ age in years $+ 7.6 \times$ waist circumference in cm $- 2.7 \times$ height in cm $+ 47 \times$ hypertension status $- 2 \times$ physical activity (at least 1 h/week) $+ 30 \times$ former smoking $+ 50 \times$ current smoking $+ 70 \times$ family history of type 2 diabetes.

We tested prediction performance of all models using the receiver operating characteristic area under the 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 0, thus performing a form of model selection. The LASSO λ was chosen by cross-validation using the squared error for Gaussian models. The GDRS was calculated in each cohort and used as a score that was fixed by setting its penalty factor

to 0 to prevent any shrinkage by LASSO. We then compared the performance of the 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.

MR

We attempted to infer causality of the replicated proteins associated with type 2 diabetes by applying two-sample bidirectional MR. Figure 1 shows the summary of the pipeline for the causal inference analysis. In summary, we extracted single nucleotide polymorphisms (SNPs) as instrumental variables (IVs) 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. (27) (N = 455,607) and the GWAS studies of SOMAscan-measured proteins by Sun et al. (28) (N = 3,301), Suhre et al. (13) (N = 1,000), and Emilsson et al. (29) (N = 5,457) for proteins. We identified ambiguous palindromic SNPs, defined as SNPs with A/T or G/C alleles and an effect allele frequency of ~0.5 using the cutoff points defined by the TwoSampleMR package in R (30). We replaced these with a proxy SNP, defined as a SNP with r^2 >0.85 with the SNP in question, when available, or excluded them from further analyses

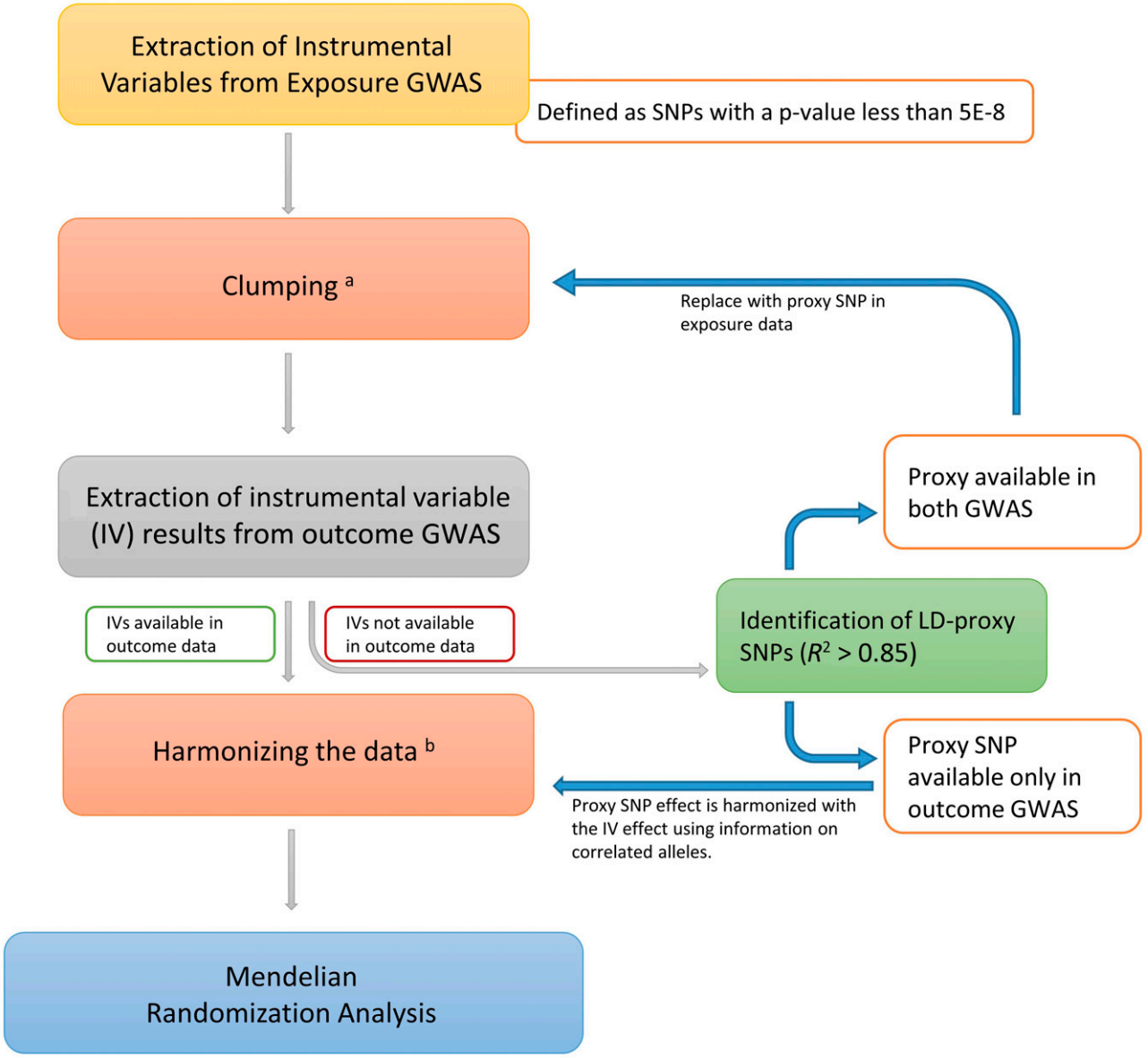

Figure 1—MR analysis flowchart. [a]Clumping refers to the process of selecting only the independent IVs (i.e., those that are not in linkage disequilibrium [LD] with one another) using the cutoff LD r^2 >0.001. [b]Harmonizing the data refers to ensuring that the effects of the IV on the exposure and the outcome reflect the same strand effect.

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/m^2), mean (SD)	27.79 (4.58)	28.36 (3.96)	0.003
Waist circumference (cm), mean (SD)	94.51 (13.81)	100.01 (11.01)	<0.001
Physical inactivity	376 (37.9)	472 (49.2)	<0.001
Smoking status			<0.001
Never	423 (42.6)	234 (24.9)	
Former	422 (42.5)	504 (53.6)	
Current	148 (14.9)	202 (21.5)	
Family history of diabetes	312 (38.1)	280 (31.6)	0.005
Hypertension	396 (39.9)	389 (41.4)	0.531

Data are mean (range) or n (%) unless otherwise indicated. *Continuous variables were tested for a difference between the two populations using t tests and categorical variables with χ^2 tests with continuity correction.

(31). We then clumped the SNPs, which implies removing SNPs in linkage disequilibrium with the lead SNP using the r^2 cutoff 0.001. We did not manually prune the final list of IVs. Furthermore, IVs selected for proteins needed to be in *cis* (i.e., within 1 Mb of the protein-coding gene as per 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) and 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 Supplementary Fig. 1. All analyses were done using R version 3.5.1 software (The R Foundation for Statistical Computing). For MR analysis, the TwoSampleMR package of R version 0.4.22 was used (30).

Data and Resource Availability

Informed consents given by KORA study participants do not cover data posting in public databases. However, the KORA data are 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 the HUNT Research Centre (https://www.ntnu.edu/hunt/data). The data used in the MR analysis are publicly available and can be accessed through https://cnsgenomics.com/content/data (Xue et al. [27]), https://www.phpc.cam.ac.uk/ceu/proteins (Sun et al. [28]), https://metabolomics.helmholtz-muenchen.de/pgwas/ (Suhre et al. [13]), and www.sciencemag.org/content/361/6404/769/suppl/DC1 (Emilsson et al. [29]). Example code for the analytic steps of the article can be accessed at https://github.com/maelhadad/T2D_SOMAscan_Proteomics.

RESULTS

Descriptive Statistics of the Study Populations

Table 1 and Supplementary Table 1 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 2), of which 24 successfully replicated in HUNT (Table 2 and Fig. 2*A*). Of these, osteomodulin was most strongly associated (on the basis of KORA P value) with an 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 (PYY) had the strongest association (1.34 [1.1–1.62] in KORA and 1.58 [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 10 FDR-significant protein associations (Supplementary Table 3). Of these, aminoacylase-1, growth hormone receptor, and IGFBP-2 replicated in HUNT (Table 3 and Fig. 2*B*). 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 1.6 [1.26–2.05] in HUNT). Interestingly, aminoacylase-1 overlapped between the replicated results of both prevalent and incident type 2 diabetes.

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

			KORA (n = 993)			HUNT (n = 940)		
Protein full name	Protein short name	UniProt identifier	OR (95% CI)	P value	FDR P value	OR (95% CI)	P value	FDR P value
α-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-β receptor type 3	TGFbR3	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	WIF1	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

Sorted alphabetically. All analyses were adjusted for age, sex, BMI, smoking status, and hypertension.

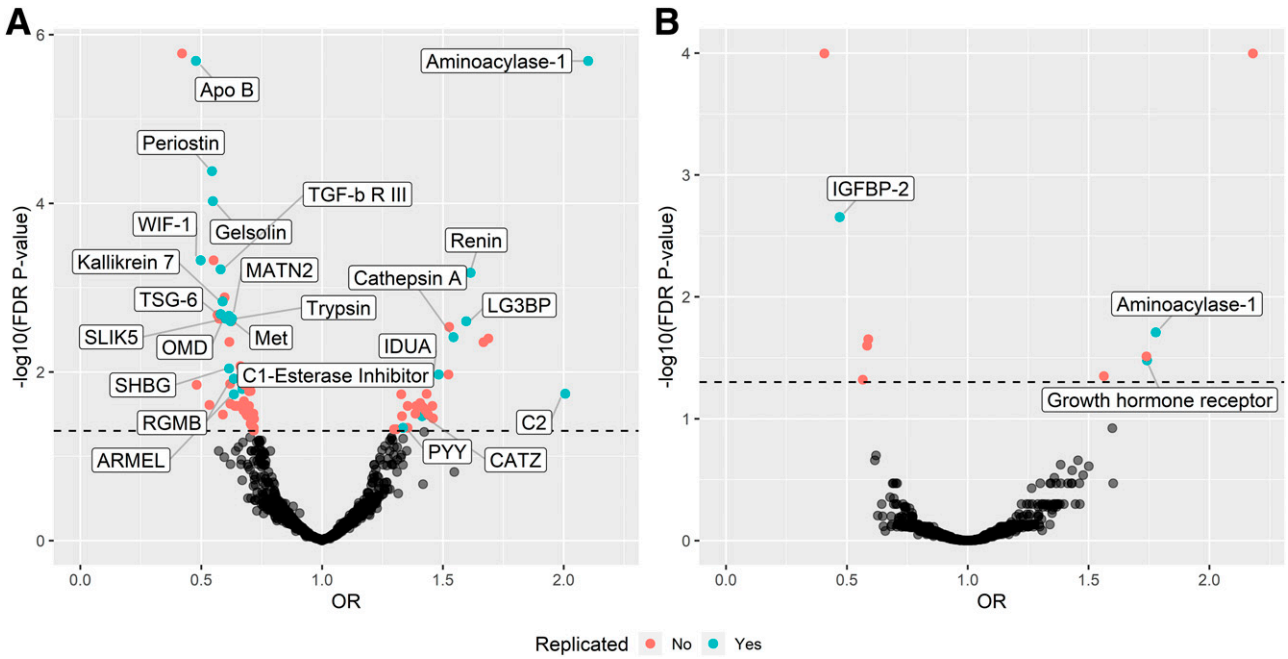

Figure 2—Volcano plot of type 2 diabetes results in KORA, where proteins that replicated in HUNT are labeled. *A*: Results of the proteome-wide analysis with prevalent type 2 diabetes in KORA. *B*: Results of the proteome-wide analysis with incident type 2 diabetes in KORA.

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

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 with protein lists of interest published by the Human Diabetes Proteome Project, namely the 1,000 diabetes-related proteins, the human islet of Langerhans proteome, the rodent β-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 4).

Data Analytics of Replicated Proteins

Supplementary Table 5 shows information extracted from Pharos for our replicated proteins. α-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 antidiabetic medication intake (Supplementary Table 6 and Supplementary Fig. 2). All the associations retained the same direction of effect apart from PYY, which showed an opposite effect after adjusting for antidiabetic 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-β 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 [95% CI 0.79–0.89] and 0.67 [0.61–0.72]; GDRS-only AUC 0.77 [0.71–0.83] and 0.66 [0.60–0.72], respectively); however, according to the Delong test, the AUC improvement in HUNT was not statistically significant ($P = 0.72$) (Supplementary Fig. 3). The calibration plot of the LASSO-selected model in HUNT yielded an intercept of 0.23 and a slope of 0.53 (Supplementary Fig. 3). The intercept of the calibration plot examines the difference of means of predicted and observed risk. In HUNT, it is >0, 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 vs. 7 years in KORA)

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

			KORA (n = 881)			HUNT (n = 794)		
Protein full name	Protein short name	UniProt identifier	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

Sorted alphabetically. All analyses were adjusted for age, sex, BMI, smoking status, and hypertension.

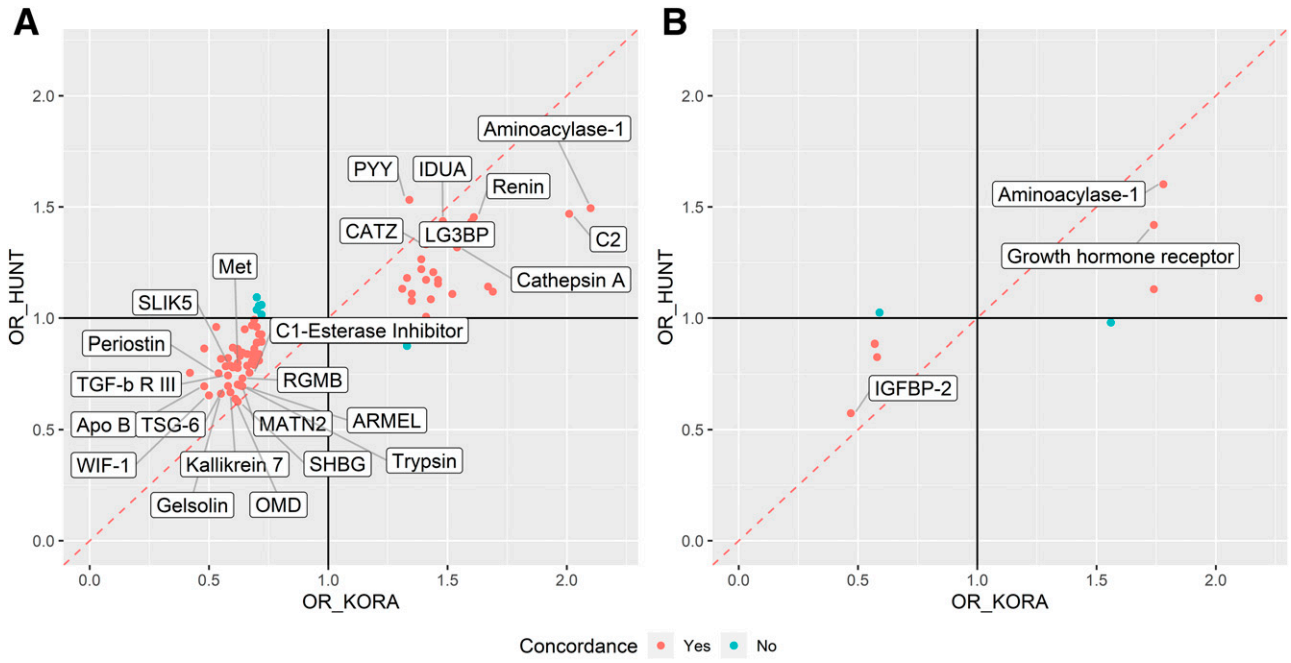

Figure 3—Coefficient concordance between KORA and HUNT for prevalent type 2 diabetes (*A*) and incident type 2 diabetes (*B*). Proteins that replicated in HUNT are labeled.

and to the fact that HUNT is older than KORA and therefore has 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 data set used gold standard screening to define type 2 diabetes, where HUNT did not apply a 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 from 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 Fig. 4). The following proteins showed relatively similar performance in both cohorts: aminoacylase-1 (KORA AUC 0.78 [95% CI 0.73–0.84]; HUNT AUC 0.71 [0.65–0.77]), growth hormone receptor (KORA AUC 0.77 [0.71–0.83]; HUNT AUC 0.70 [0.64–0.76]), and IGFBP-2 (KORA AUC 0.78 [0.72–0.84]; HUNT AUC 0.73 [0.68–0.79]).

MR Analysis of Replicated Plasma Proteins and Type 2 Diabetes in KORA

Using up to 120 SNPs as genetic instruments, we investigated whether type 2 diabetes had a causal effect on the 26 replicated proteins from both the prevalence and the incidence analyses (Supplementary Table 7 and Supplementary Fig. 5). For cathepsin Z (MR inverse variance–weighted $\beta = 0.13$; $P = 2.00e\text{-}03$) and renin (0.08; $P = 3.15e\text{-}02$), a nominally significant causal effect of prevalent type 2 diabetes was observed, each with the same direction of effect as its observational results. MR-Egger analyses to test for the presence of horizontal pleiotropy showed no significant results for either protein (intercept $P = 0.17$ and 0.1 for cathepsin Z and renin, respectively). Tests and plots to check for outliers in the IVs showed no significant aberrations (Supplementary Figs. 6 and 7).

We also ran MR to investigate whether 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 8 and Supplementary Fig. 5). 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 $\beta = -0.09$; $P = 2.95e\text{-}02$). 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, 2 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 downregulated 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 patients with type 2 diabetes (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 IGF-I 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 and NTRK-like protein 5, which were downregulated in type 2 diabetes cases, and lysosomal protective protein, galectin-3 binding protein (LGALS3BP), and 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 bypass 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 nonalcoholic 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, respectively, in our results, 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 of 85 for prevalent and 8 of 10 for incident type 2 diabetes) showed concordant effects between cohorts, while none of the nonconcordant proteins were statistically significant in the replication (Fig. 3). Nonetheless, fasting has significant metabolic consequences that are expected to be reflected in the plasma proteome and could have contributed to nonreplication in HUNT. However, there are multiple other potential explanations for the nonreplication, perhaps differing from one protein to another. Importantly, while plasma protein levels differ between fasting and nonfasting 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 soluble CD163 (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 that fasting status may not be the primary reason for nonreplication 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 antidiabetic 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, because 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 ($P = 0.72$) 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 IGFBP-2, each of which achieved approximately equal performance in HUNT compared with KORA, results that warrant validation in clinical trials using commercially available ELISA kits. Because the KORA samples were taken from individuals in a fasting state ($\geq$8 h) and HUNT samples were taken nonfasting, 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 with KORA for some of the other measured biomarkers in combination with the GDRS.

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 cardiovascular disease 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 β-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 publicly 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 nonspecific binding (28,29). However, we verified that none of the proteins identified have been flagged for such issues (validation data presented in Supplementary Material, Supplementary Table 10, and Supplementary Figs. 10 and 11). Because of the lack of oral glucose tolerance test data in HUNT, the rigorous definition of type 2 diabetes used 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 MR 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 data sets 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 aminoacylase-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 believed to play a role in the development of cardiovascular complications, like atherosclerosis. MR 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 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.

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 and by the State of Bavaria. This work was supported by Deutsche Forschungsgemeinschaft grant WA 4081/1-1 and by the Bundesministerium für Bildung und Forschung within the framework of the European Union 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, Norges Teknisk-Naturvitenskapelige Universitet, Norges Forskningsråd, Helse Midt-Norge, the Nord-Trøndelag County Council, and the Norwegian Institute of Public Health. M.A.E. is supported by the Deutsches Zentrum für Herz-Kreislaufforschung. T.I.O. is supported by National Institutes of Health grants U24-CA-224370, U24-TR-002278, and U01-CA-239108. K.S. is supported by Qatar National Research Fund grant NPRPC11-0115-180010.

Duality of Interest. C.J. has received personal fees for research consultancy work from Pfizer and Bayer outside 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 the submitted work. No other potential conflicts of interest relevant to this article were reported.

Author Contributions. M.A.E. analyzed the data, interpreted the results, and wrote and revised the manuscript. M.A.E. and M.W. designed the study. R.W., V.G.-D., and T.I.O. helped with the analyses. C.J., C.H., R.W., C.G., P.M., H.G., J.G., W.R., C.v.T., S.M.H., W.K., M.F.S., K.S., B.T., K.H., A.P., and M.W. were involved in the data collection, data management, and preparation of their respective cohorts. All authors contributed to the writing of the article, 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.

Prior Presentation. Parts of this study were presented in abstract form at the AHA Epidemiology and Prevention–Lifestyle and Cardiometabolic Health 2020 Scientific Sessions, Phoenix, AZ, 3–6 March 2020. This study has been published previously in abstract form in the following publication: Elhadad MA, Jonasson C, Huth C, et al. Deciphering the plasma proteome of type 2 diabetes. Circulation 2020;141(Suppl. 1):A21.

References

1. International Diabetes Federation. IDF Diabetes Atlas, 9th edition, 2019. Brussels, Belgium, International Diabetes Federation. Accessed 23 March 2020. Available from https://www.diabetesatlas.org

2. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol 2018;14:88–98

3. Huth C, von Toerne C, Schederecker F, et al. Protein markers and risk of type 2 diabetes and prediabetes: a targeted proteomics approach in the KORA F4/FF4 study. Eur J Epidemiol 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;62:1998–2006

5. Nowak C, Sundström J, Gustafsson S, et al. 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, et al. 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. Endocr Connect 2015;4:249–260
8. Zheng J, Baird D, Borges MC, et al. Recent developments in Mendelian randomization studies. Curr Epidemiol Rep 2017;4:330–345
9. Paprott R, Mühlenbruch K, Mensink GBM, et al. Validation of the German Diabetes Risk Score among the general adult population: findings from the German Health Interview and Examination Surveys. BMJ Open Diabetes Res Care 2016;4: e000280
10. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 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, Löwel H, Wichmann HE; MONICA/KORA Study Group. KORA–a research platform for population based health research. Gesundheitswesen 2005;67(Suppl. 1):S19–S25
13. Suhre K, Arnold M, Bhagwat AM, et al. Connecting genetic risk to disease end points through the human blood plasma proteome [published correction appears in Nat Commun 2017;8:15345]. Nat Commun 2017;8:14357
14. Krokstad S, Langhammer A, Hveem K, et al. Cohort profile: the HUNT study, Norway. Int J Epidemiol 2013;42:968–977
15. Ganz P, Heidecker B, Hveem K, et al. 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, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. PLoS One 2010;5:e15004
17. Mühlberger N, Behrend C, Stark RG, Holle R. Datenbankgestutzte Online-Erfassung von Arzneimitteln im Rahmen gesundheitswissenschaftlicher Studien - Erfahrungen mit der IDOM-Software. München, Germany, Urban & Fischer, 2003
18. Nguyen DT, Mathias S, Bologa C, et al. Pharos: collating protein information to shed light on the druggable genome. Nucleic Acids Res 2017;45:D995–D1002
19. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res 2019;47:D330–D338
20. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. Nucleic Acids Res 2019;47:D590–D595
21. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 2019;47:D1005–D1012
22. Amberger JS, Bocchini CA, Scott AF, Hamosh A. OMIM.org: leveraging knowledge across phenotype-gene relationships. Nucleic Acids Res 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, et al. The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. Nucleic Acids Res 2014;42:D802–D809
25. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc B 1996;58:267–288
26. Steyerberg EW, Vickers AJ, Cook NR, et al. 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, et al.; eQTLGen Consortium. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. Nat Commun 2018;9:2941
28. Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. Nature 2018;558:73–79
29. Emilsson V, Ilkov M, Lamb JR, et al. Co-regulatory networks of human serum proteins link genetics to disease. Science 2018;361:769–773
30. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife 2018;7:e34408
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. Int J Epidemiol 2016;45:1717–1726
32. Teumer A. Common methods for performing Mendelian randomization. Front Cardiovasc Med 2018;5:51
33. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol 2015;44:512–525
34. Topf F, Schvartz D, Gaudet P, et al. The Human Diabetes Proteome Project (HDPP): from network biology to targets for therapies and prevention. Transl Proteom 2013;1:3–11
35. Pérez-Pérez R, García-Santos E, Ortega-Delgado FJ, et al. Attenuated metabolism is a hallmark of obesity as revealed by comparative proteomic analysis of human omental adipose tissue. J Proteomics 2012;75:783–795
36. Caira S, Iannelli A, Sciarrillo R, et al. Differential representation of liver proteins in obese human subjects suggests novel biomarkers and promising targets for drug development in obesity. J Enzyme Inhib Med Chem 2017;32:672–682
37. Cao Y-F, Li J, Zhang Z, et al. Plasma levels of amino acids related to urea cycle and risk of type 2 diabetes mellitus in Chinese adults. Front Endocrinol (Lausanne) 2019;10:50
38. Hedbacker K, Birsoy K, Wysocki RW, et al. Antidiabetic effects of IGFBP2, a leptin-regulated gene. Cell Metab 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, et al. Plasma gelsolin levels decrease in diabetic state and increase upon treatment with F-actin depolymerizing versions of gelsolin. J Diabetes Res 2014;2014:152075
41. Joseph JJ, Echouffo Tcheugui JB, Effoe VS, Hsueh WA, Allison MA, Golden SH. Renin-angiotensin-aldosterone system, glucose metabolism and incident type 2 diabetes mellitus: MESA. J Am Heart Assoc 2018;7:e009890
42. Le TN, Nestler JE, Strauss JF III, Wickham EP III. Sex hormone-binding globulin and type 2 diabetes mellitus. Trends Endocrinol Metab 2012;23:32–40
43. Geyer PE, Wewer Albrechtsen NJ, Tyanova S, et al. 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, et al. Plasma proteome profiling reveals dynamics of inflammatory and lipid homeostasis markers after Roux-en-Y gastric bypass surgery. Cell Syst 2018;7:601–612.e3
45. Niu L, Geyer PE, Wewer Albrechtsen NJ, et al. 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, et al. 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. Proc Natl Acad Sci U S A 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, et al. 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, et al. 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, et al. Transforming growth factor-β receptor III is a potential regulator of ischemia-induced cardiomyocyte apoptosis. J Am Heart Assoc 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, et al. 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, Gårdinger Y, Björgell O, Hlebowicz J. Effect of food intake on biomarkers for cardiovascular disease and inflammation analyzed with the Proseek Multiplex CVD II kit. Genet Mol Res 2018;17:gmr16039884

56. Bahtiyar G, Gutterman D, Lebovitz H. Heart failure: a major cardiovascular complication of diabetes mellitus. Curr Diab Rep 2016;16:116

57. Pasquier A, Vivot K, Erbs E, et al. Lysosomal degradation of newly formed insulin granules contributes to β cell failure in diabetes. Nat Commun 2019;10: 3312

4. Paper II:

Metabolic syndrome and the plasma proteome: from association to causation

Mohamed A. Elhadad, Rory Wilson, Shaza B. Zaghlool, Cornelia Huth, Christian Gieger, Harald Grallert, Johannes Graumann, Wolfgang Rathmann, Wolfgang Koenig, Moritz F. Sinner, Kristian Hveem, Karsten Suhre, Barbara Thorand, Christian Jonasson, Melanie Waldenberger & Annette Peters

Cardiovasc Diabetol 20, 111 (2021)

https://doi.org/10.1186/s12933-021-01299-2

Elhadad *et al. Cardiovasc Diabetol* (2021) 20:111
https://doi.org/10.1186/s12933-021-01299-2

Cardiovascular Diabetology

ORIGINAL INVESTIGATION Open Access

Metabolic syndrome and the plasma proteome: from association to causation

Mohamed A. Elhadad[1,2,3*], Rory Wilson[1,2], Shaza B. Zaghlool[4], Cornelia Huth[2,5], Christian Gieger[1,2,5], Harald Grallert[1,2,5], Johannes Graumann[6,7], Wolfgang Rathmann[5,8], Wolfgang Koenig[3,9,10], Moritz F. Sinner[3,11], Kristian Hveem[12,13], Karsten Suhre[4], Barbara Thorand[2,5], Christian Jonasson[12,13], Melanie Waldenberger[1,2,3*] and Annette Peters[2,3,5,14*]

Abstract

Background: The metabolic syndrome (MetS), defined by the simultaneous clustering of cardio-metabolic risk factors, is a significant worldwide public health burden with an estimated 25% prevalence worldwide. The pathogenesis of MetS is not entirely clear and the use of molecular level data could help uncover common pathogenic pathways behind the observed clustering.

Methods: Using a highly multiplexed aptamer-based affinity proteomics platform, we examined associations between plasma proteins and prevalent and incident MetS in the KORA cohort (n = 998) and replicated our results for prevalent MetS in the HUNT3 study (n = 923). We applied logistic regression models adjusted for age, sex, smoking status, and physical activity.

We used the bootstrap ranking algorithm of least absolute shrinkage and selection operator (LASSO) to select a predictive model from the incident MetS associated proteins and used area under the curve (AUC) to assess its performance. Finally, we investigated the causal effect of the replicated proteins on MetS using two-sample Mendelian randomization.

Results: Prevalent MetS was associated with 116 proteins, of which 53 replicated in HUNT. These included previously reported proteins like leptin, and new proteins like NTR domain-containing protein 2 and endoplasmic reticulum protein 29. Incident MetS was associated with 14 proteins in KORA, of which 13 overlap the prevalent MetS associated proteins with soluble advanced glycosylation end product-specific receptor (sRAGE) being unique to incident MetS. The LASSO selected an eight-protein predictive model with an (AUC = 0.75; 95% CI = 0.71–0.79) in KORA.

Mendelian randomization suggested causal effects of three proteins on MetS, namely apolipoprotein E2 (APOE2) (Wald-Ratio = − 0.12, Wald-p = 3.63e−13), apolipoprotein B (APOB) (Wald-Ratio = − 0.09, Wald-p = 2.54e−04) and proto-oncogene tyrosine-protein kinase receptor (RET) (Wald-Ratio = 0.10, Wald-p = 5.40e−04).

Conclusions: Our findings offer new insights into the plasma proteome underlying MetS and identify new protein associations. We reveal possible casual effects of APOE2, APOB and RET on MetS. Our results highlight protein candidates that could potentially serve as targets for prevention and therapy.

*Correspondence: mohamed.elhadad@helmholtz-muenchen.de; waldenberger@helmholtz-muenchen.de; peters@helmholtz-muenchen.de
[2] Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
Full list of author information is available at the end of the article

Keywords: Metabolic syndrome, Proteomics, Blood proteins, Mendelian randomization analysis, Diabetes mellitus, type 2, Cardiovascular disease, Risk factors

Background

The metabolic syndrome (MetS) is a constellation of risk factors significantly increasing the risk of type 2 diabetes (T2D) and cardiovascular diseases (CVD) like coronary artery disease (CAD), stroke and heart failure [1, 2]. The respective risk factors are increased waist circumference, hypertriglyceridemia, reduced high-density lipoprotein, hyperglycemia and increased blood pressure. The prevalence of MetS has been steadily increasing in recent decades in conjunction with the obesity pandemic, driven by surplus eating and a sedentary lifestyle [3, 4]. It is estimated that 25% of adults worldwide have MetS, causing significant financial impact on healthcare systems [5].

Since its conception, the nature of MetS has been under debate [6–9]. However, most researchers agree that the clustering of the above mentioned risk factors is more frequent than could be attributed to chance alone [6–9]. In the center of the debate is MetS' pathogenesis, which remains in the hypothesis stage. Suggested common driving pathogenic pathways include visceral adiposity and insulin resistance with subsequent dyslipidemia and subclinical inflammation [6]. While the suggested pathways help partly explain the clustering of risk factors and increased risk in some patients, they fail to explain the lack or incomplete clustering of those risk factors in others.

Recently, the introduction of omics data into CVD research has helped uncover molecular pathophysiological players, an example being the identification of PCSK9 as a drug target through genetic studies of CAD [10]. Omics studies with regard to cardio-metabolic risk factors have also been informative. Using the UK-Biobank data, a recent genetic study of MetS identified loci that are common to all MetS components as well as loci that are unique to the syndrome, i.e. not associated with the components themselves [11].

Proteomics, the study of proteins, can provide insight into the downstream players of genetics in the molecular pathogenic pathway of MetS and identify predictive biomarkers or targets for drug development. Enabled by advances in proteomics, studies with MetS have expanded from single protein to multi-protein investigations, the largest to date featuring 249 proteins [12]. Reported protein associations with MetS include adipokines like leptin and adiponectin (ADIPOQ), liver secreted proteins like sex hormone binding globulin (SHBG) and inflammatory markers like C-reactive protein, tumor necrosis factor alpha and complement system proteins [13, 14]. These proteins indicate functional links to MetS-defining features such as insulin resistance and visceral adiposity, and help explain the increased risk of complications, like CVD, in MetS patients.

In the present study, we use a highly multiplexed, aptamer-based, affinity proteomics platform (SOMAscan™) to assess the association between 1095 blood plasma proteins and prevalent and incident MetS in the KORA cohort, and replicate our results in the HUNT study. The proteins assessed by the SOMAscan platform have been selected to represent markers of a broad range of biological pathways and tissue specific processes. We investigate the use of these proteins as biomarkers and explore their potential causal effects using two-sample Mendelian randomization (MR) [15].

Methods

Study populations

KORA cohort

The KORA study (Cooperative health research in the Region of Augsburg) is a population-based cohort study from Augsburg, southern Germany. The study was approved by the ethics committee of the Bavarian Medical Association. Written informed consent was obtained from each participant. We used KORA-F4 (conducted 2006–2008) for cross-sectional analysis of prevalent MetS and its follow-up survey KORA-FF4 (conducted 2013–2014) for the prospective analysis of incident MetS (mean follow-up time = 6.5, SD = 0.5 years). For both surveys, detailed clinical and demographic information was collected, as was peripheral blood for later omics analyses. Details on the KORA cohort have been previously published [16]. A random subsample of 1000 individuals was selected from the already deeply phenotyped KORA-F4 study participants for proteomics measurement using the SOMAscan platform. One sample was excluded because it failed SOMALogic quality control and one participant was excluded due to the lack of sufficient information to define MetS leaving 998 participants for the final cross-sectional analysis. For the follow-up analysis, 371 participants with prevalent MetS and four participants lacking sufficient information to define incident MetS were excluded leaving 623 participants for analysis.

HUNT cohort

The Nord-Trøndelag Health Study (HUNT) is a population-based cohort from Nord-Trøndelag County in

Norway. We used the HUNT3 survey (performed 2006–2008, N = 1017 with proteomics measurements) for the replication of the KORA study cross-sectional results. The HUNT study collected detailed socio-demographic and clinical information for all participants [17]. Ten samples failed SOMALogic quality control and were excluded from further analyses. Moreover, fourteen participants were excluded due to a lack of sufficient information to define MetS and an additional 70 participants were excluded due to missing information for the covariates smoking status and physical activity, leaving 923 participants for the final cross-sectional analysis.

Proteomics measurement

The aptamer based SOMAscan platform was used to quantify proteins in both cohorts. Details on the platform [18] and its application to the KORA cohort have been described before [19]. In brief, each aptamer was selected to have high affinity toward a specific protein. Plasma was incubated with the aptamer mix and then exposed to multiple washing steps in the form of 2 bead-based immobilization steps to eliminate unbound or unspecifically bound aptamers and proteins. Finally, aptamers were eluted from the proteins and quantified as proxies to protein concentration by hybridization to custom arrays of aptamer-complementary oligonucleotides. The resulting raw intensities were processed with the help of standard samples included on each plate using a data analysis workflow consisting of hybridization normalization, median signal normalization and signal calibration to control for inter-plate differences [18]. The raw intensities are reported as relative florescence units.

Fasting plasma samples from the KORA study were sent to SomaLogic Inc. (Boulder Colorado, USA) for analysis [19]. Of the 1129 SOMAmer aptamers (SOMAscan assay V3.2) 29 failed SOMAscan quality control. We additionally removed five aptamers as recommended by the SOMAscan assay change log issued on December 22, 2016, leaving 1095 aptamers for analysis. For replication, we used only the HUNT aptamers that passed quality control [20].

MetS definition in KORA

MetS was defined according to the harmonized definition by Alberti et al. [21] by the presence of three or more of the following criteria: (1) waist circumference ≥ 94 cm in men or ≥ 80 cm in women; (2) fasting serum triglycerides ≥ 150 mg/dl or drug treatment for elevated triglycerides (fibrates); (3) serum high density lipoprotein cholesterol (HDL) < 40 mg/dl in men or < 50 mg/dl in women or drug treatment for reduced HDL (fibrates); (4) systolic blood pressure ≥ 130 mmHg or diastolic blood pressure ≥ 85 mmHg or treatment with antihypertensive medication; (5) fasting serum glucose level ≥ 100 mg/dl or intake of antidiabetic medication.

MetS definition in HUNT

The same definition was used for HUNT with some differences due to the unavailability of fasting measurements and information on drug treatment for elevated triglycerides or reduced HDL. For defining lipid components, we applied the cut-off levels suggested by Driver et al. for the diagnosis of metabolic syndrome using non-fasting lipid measurements [22]. For defining the low HDL component of MetS, we applied the same cut-off levels as for KORA [22]. For defining high triglycerides, we used a cut-off of 200 mg/dl [22]. For defining the hyperglycemia component, we used a cut-off of 140 mg/dl suggested by the American diabetes association diabetes diagnosis guideline to diagnose impaired glucose tolerance [23] or intake of antidiabetic medication.

Statistical analysis

SOMAscan data was log2 transformed and each protein was standardized to have a mean of zero and a SD of 1 by subtracting its mean and dividing by its standard deviation to allow easier interpretation of the results per SD of log-transformed protein level.

Baseline characteristics were compared between the two cohorts using t-tests for continuous variables and chi-square tests with continuity correction for categorical variables.

Proteome-wide analysis

Proteome-wide analyses to test for associations between prevalent and incident MetS and proteins were carried out using logistic regression with one model per protein. Each model had prevalent or incident MetS as the outcome, the log-transformed protein level as the explanatory variable, and was adjusted for age, sex, smoking status (categorized as never smoker, former smoker and current smoker) and physical activity (categorized as active vs inactive). We applied the Bonferroni method to correct for multiple testing throughout the paper. For the proteome-wide analyses this resulted in a significance threshold of $p < 4.6e{-}05$ (0.05/1095).

To replicate our results for prevalent MetS, we applied the same model in HUNT. We considered results replicated if they had consistent effect direction and survived Bonferroni correction calculated based on the number of KORA significant proteins.

Furthermore, we assessed the association of individual prevalent and incident MetS components with replicated prevalent MetS proteins and KORA incident MetS significant results, respectively. For incident components, analysis was done after removing participants with MetS

at baseline. For each component (increased waist circumference, hypertriglyceridemia, reduced HDL, hyperglycemia and increased blood pressure), we applied the same model with the component as an outcome using the KORA data.

Biomarker discovery for MetS

We investigated the predictive utility of the proteins significantly associated with incident MetS in KORA by utilizing the bootstrap ranking algorithm of the least absolute shrinkage and selection operator (LASSO) for model selection using the "elasso" R-package version 1.1 [24]. LASSO attempts to shrink the coefficients of the model covariates to zero thus selecting the covariates with the best predictive ability. We applied cross-validation to select the best LASSO constraint "lambda" within each bootstrap iteration.

We then used the area under the receiver operating characteristic curve (ROC-AUC) to test model performance calculated using the "pROC" R-package version 1.16.2 [25]. We further assessed performance using the calibration plot, which examines the agreement between observed and fitted values of the outcome [26] and by comparing the performance of LASSO selected protein model to a baseline model based on age and sex utilizing the DeLong test [27].

Additionally, we tested the performance of proteins associated with prevalent MetS as a biomarker panel using KORA as a training dataset and HUNT as a test dataset (full details in Additional file 1).

Enrichment and protein–protein interaction network analyses

We used STRING [28] to evaluate the protein–protein interaction network of the MetS associated proteins (full details in Additional file 1).

Mendelian randomization analysis

We used two-sample MR to investigate potential causal effects of replicated proteins on MetS. Mendelian randomization analysis is an instrumental variable (IV) analysis, in which genetic associations are used as anchors to assess causal effects of an exposure of interest on an outcome of interest. Two-sample MR entails the use of published genetic; i.e. single nucleotide polymorphism (SNP) association results to obtain IVs, thus allowing the use of the available bigger sample sizes and meta-analyses of genome wide association studies (GWAS).

First, we extracted SNPs associated with the protein of interest from already published genetic association studies using data of European ancestry. We extracted the IVs from SOMAscan GWAS studies by Suhre et al. (n = 1000) [19] and Sun BB et al. (n = 3301) [29] and the cis only association study by Emilsson et al. (n = 5457) [30].

We then identified ambiguous palindromic SNPs, which are SNPs with A/T or G/C alleles and an effect allele frequency around 0.5, using the cut-off points defined by the "TwoSampleMR" R-package [15]. We replaced the SNPs in question with an available proxy, defined as a SNP with r2 exceeding 0.85, or excluded them from further analyses [31]. To obtain a list of independent SNPs to be used as IVs in further analyses, we clumped the list of SNPs using the r2 cut-off of 0.001. Selected IVs had to be in cis with the protein of interest, i.e., within one Mb of the protein-coding gene as per the Human Genome Assembly GRCh37.p13. We subsequently extracted the outcome summary statistics of the selected IVs or of one of their proxies from the MetS GWAS study by Lind (n = 291,107) [11].

We used the Wald ratio to estimate a causal effect if there was only one IV available [32]. In cases where more IVs were available, we applied a random effects model of the inverse variance weighted meta-analysis to combine the Wald ratio estimates of all IVs [32, 33]. Whenever there was more than one IV, we ran the MR-Egger regression model to check for horizontal pleiotropy in our causal models [34], and we investigated scatter plots, leave-one-out analysis plots and forest plots to identify outliers among the IVs that could be driving the results in a certain direction.

All analyses were done in R version 4.0.2 (The R Foundation for Statistical Computing). For MR analysis, the "TwoSampleMR" R-package version 0.5.5 was employed [15].

Results

Descriptive statistics of the study populations

Table 1 shows the baseline characteristics of both cohorts. The KORA sample comprised 998 participants with an age range of 43–75 years, of whom 515 were women, 371 had MetS at baseline and 147 developed it between baseline and follow-up. The HUNT sample compromised 923 participants with an age range of 31.6–91.7 years, of whom 235 were women and 418 had MetS. KORA participants had significantly lower waist circumference and triglyceride levels, and higher HDL levels, and were less often current smokers. Baseline characteristics of the follow-up subset of KORA used in incident MetS analyses are shown in Table 1.

Association results of plasma proteins with prevalent MetS

The proteome-wide analysis of prevalent MetS yielded 116 Bonferroni significant proteins, of which 51 are positively associated with MetS and 65 are negatively associated (Additional file 2: Table S1). Of these, 53 successfully

Table 1 Baseline characteristics of the study populations

Variable	Prevalent MetS			Incident MetS
	KORA (n = 998)	HUNT (n = 923)	p value*	KORA (n = 623)
Age[a] (years)	59.3 (43–75)	68.93 (31.6–91.7)	< 0.001	58.15 (43–74)
Sex female[b]	515 (51.6%)	235 (25.5%)	< 0.001	379 (60.8%)
BMI[c] (kg/m^2)	27.77 (4.58)	28.39 (3.97)	0.002	26.21 (3.87)
Waist circumference[c] (cm)	94.56 (14.05)	100.18 (11.04)	< 0.001	89.07 (11.44)
Waist hip ratio[c]	0.89 (0.08)	0.96 (0.07)	< 0.001	0.86 (0.08)
Physically active[b]	620 (62.1%)	468 (50.7%)	< 0.001	416 (66.8%)
Smoking[b]			< 0.001	
Never smoker	423 (42.4%)	231 (25%)		277 (44.5%)
Former smoker	427 (42.8%)	497 (53.8%)		244 (39.2%)
Current smoker	148 (14.8%)	195 (21.1%)		102 (16.4%)
Total cholesterol[c] (mg/dl)	221.99 (38.47)	178.50 (42.34)	NA**	222.83 (37.11)
HDL- cholesterol[c] (mg/dl)	57.35 (15.19)	45.05 (11.25)	NA**	62.84 (14.45)
Triglyceride level[c] (mg/dl)	129.06 (87.68)	161.82 (86.81)	NA**	96.59 (46.08)
Hypertension[d]	398 (39.9%)	382 (41.4%)	0.544	143 (23.0%)

* Continuous variables were tested for a difference between the two populations using t-tests and categorical variables with chi-square tests with continuity correction, **Differences between cohorts could not be statistically tested as KORA was measured in fasting samples and HUNT in non-fasting samples

[a] Mean (range)

[b] Number (percentage)

[c] Mean ± standard deviation. Hypertension was defined as having systolic blood pressure ≥ 140 mmHg and diastolic ≥ 90 mmHg or known medication-controlled hypertension. In HUNT we additionally used the ICD-10 codes I10–I15 of the hospital and primary care data and the codes K86 or K87 of the International Classification of Primary Care, Second Edition, to identify participants with hypertension

replicated in HUNT (Table 2; Fig. 1a). All of the 56 non-replicated proteins available in HUNT showed concordant direction of effect between the cohorts, and 35 of them were nominally significant in HUNT (Fig. 1b).

Among the replicated proteins, insulin-like growth factor-binding protein 2 (IGFBP2) had the lowest odds ratio (OR) in both cohorts per SD increase in log-transformed protein level, with values of 0.33 (95% CI 0.27–0.39) in KORA and of 0.52 (95% CI 0.44–0.62) in HUNT; and leptin had the highest OR in both cohorts, with values of 3.7 (95% CI 2.95–4.7) in KORA and 1.76 (95% CI 1.49–2.08) in HUNT. The correlation matrices of replicated proteins are shown in Additional file 1: Figure S1.

Association results of plasma proteins with incident MetS

The proteome-wide analysis of incident MetS in KORA yielded 14 significant protein associations at a Bonferroni corrected threshold (Table 3; Fig. 2). IGFBP2 was the most strongly associated protein based on p-value (OR = 0.55; 95% CI = 0.44–0.68) and plasminogen activator inhibitor 1 (SERPINE1) had the largest magnitude of association (OR = 3.70; 95% CI = 2.95–4.70). The incident MetS significant proteins included 10 overlapping the replicated results and 13 overlapping the KORA significant results of prevalent MetS (Fig. 2b). Only soluble advanced glycosylation end product-specific receptor (sRAGE) (OR = 0. 63; 95% CI = 0.51–0.77) was unique to incident MetS.

MetS components analysis

For each prevalent component, we tested whether the prevalent-MetS-replicated proteins were also associated with the component. Each component was associated with at least 33 of these proteins, with increased waist circumference and hypertriglyceridemia showing the highest number of associations with 50 and 48, respectively (Fig. 3; Additional file 2: Table S2). In total, 18 proteins were common to all prevalent components (Additional file 2: Table S2).

For the incident components, increased waist circumference and high blood pressure were associated with 13 and 8 proteins out of the 14 incident MetS KORA protein associations respectively (Additional file 2: Table S3). ADIPOQ and IGFBP2 were associated with the four incident components with significant results namely increased waist circumference, hypertriglyceridemia, hyperglycemia and increased blood pressure.

Biomarker discovery

We explored the utility of proteins associated with incident MetS as predictive biomarkers in KORA. The LASSO-selected predictive model included 8 proteins (Additional file 2: Table S4) and had an AUC of 0.75

Table 2 Replicated results of the proteome-wide analysis of prevalent MetS in KORA and HUNT, sorted by the magnitude of the OR in KORA

Protein full name	UniProt	Gene symbol	KORA (n = 998)		HUNT (n = 923)	
			OR (95% CI)	P-value	OR (95% CI)	P-value
Leptin	P41159	LEP	3.70 (2.95–4.70)	4.77E-28	1.76 (1.49–2.08)	3.76E−11
Plasminogen activator inhibitor 1	P05121	SERPINE1	2.51 (2.12–3.00)	2.24E−25	1.37 (1.19–1.57)	9.99E−06
Growth hormone receptor	P10912	GHR	2.33 (1.97–2.78)	3.84E−22	2.08 (1.76–2.47)	9.44E−18
Tissue-type plasminogen activator	P00750	PLAT	2.17 (1.82–2.61)	1.54E−17	1.36 (1.19–1.57)	8.57E−06
Aminoacylase-1	Q03154	ACY1	2.16 (1.83–2.56)	2.95E−19	1.93 (1.64–2.30)	2.71E−14
Dickkopf-like protein 1	Q9UK85	DKKL1	1.96 (1.63–2.39)	6.02E−12	1.31 (1.15–1.51)	9.54E−05
Galectin-3-binding protein	Q08380	LGALS3BP	1.85 (1.59–2.16)	3.66E−15	1.52 (1.32–1.75)	6.81E−09
GDNF family receptor alpha-1	P56159	GFRA1	1.78 (1.52–2.09)	7.06E−13	1.46 (1.27–1.70)	1.99E−07
Complement factor H	P08603	CFH	1.76 (1.51–2.06)	1.44E−12	1.67 (1.42–1.98)	7.71E−10
Apolipoprotein E (isoform E3)	P02649	APOE	1.73 (1.49–2.02)	1.17E−12	1.47 (1.26–1.72)	1.16E−06
Retinoic acid receptor responder protein 2	Q99969	RARRES2	1.72 (1.48–2.01)	3.04E−12	1.54 (1.33–1.80)	1.84E−08
Endoplasmic reticulum resident protein 29	P30040	ERP29	1.71 (1.46–2.00)	1.39E−11	1.40 (1.21–1.61)	3.97E−06
Complement factor I	P05156	CFI	1.70 (1.46–2.00)	1.81E−11	1.36 (1.18–1.58)	2.72E−05
Proto-oncogene tyrosine-protein kinase receptor	P07949	RET	1.70 (1.46–1.99)	1.19E−11	1.87 (1.59–2.21)	1.71E−13
Bone morphogenetic protein 1	P13497	BMP1	1.68 (1.44–1.97)	8.25E−11	1.40 (1.17–1.68)	2.66E−04
Afamin	P43652	AFM	1.66 (1.43–1.93)	4.89E−11	1.37 (1.19–1.58)	1.28E−05
Reticulon-4 receptor	Q9BZR6	RTN4R	1.65 (1.42–1.92)	7.45E−11	1.65 (1.42–1.92)	6.57E−11
Scavenger receptor cysteine-rich type 1 protein M130	Q86VB7	CD163	1.59 (1.38–1.85)	5.75E−10	1.41 (1.23–1.62)	8.61E−07
C–C motif chemokine 25	O15444	CCL25	1.57 (1.36–1.82)	1.04E−09	1.33 (1.15–1.53)	7.50E−05
E-selectin	P16581	SELE	1.56 (1.34–1.82)	1.29E−08	1.68 (1.45–1.94)	3.29E−12
Ficolin-3	O75636	FCN3	1.52 (1.30–1.77)	1.15E−07	1.31 (1.13–1.52)	3.72E−04
Lysosomal protective protein	P10619	CTSA	1.51 (1.31–1.76)	4.55E−08	1.29 (1.12–1.49)	4.00E−04
Cathepsin Z	Q9UBR2	CTSZ	1.49 (1.29–1.73)	9.24E−08	1.42 (1.24–1.65)	1.35E−06
Thrombospondin-2	P35442	THBS2	1.43 (1.24–1.65)	7.57E−07	1.38 (1.20–1.60)	1.33E−05
C–C motif chemokine 16	O15467	CCL16	1.42 (1.21–1.68)	3.69E−05	1.34 (1.16–1.55)	5.95E−05
Insulin-like growth factor-binding protein 2	P18065	IGFBP2	0.33 (0.27–0.39)	4.63E−33	0.52 (0.44–0.62)	4.04E−14
Sex hormone-binding globulin	P04278	SHBG	0.42 (0.35–0.50)	6.47E−22	0.47 (0.40–0.56)	2.86E−18
Insulin-like growth factor-binding protein 1	P08833	IGFBP1	0.43 (0.37–0.51)	1.03E−23	0.55 (0.47–0.64)	1.98E−14
Endothelial cell-specific molecule 1	Q9NQ30	ESM1	0.47 (0.39–0.56)	1.20E−16	0.68 (0.57–0.80)	1.31E−05
Netrin receptor UNC5D	Q6UXZ4	UNC5D	0.48 (0.40–0.56)	8.72E−18	0.75 (0.64–0.87)	1.92E−04
WAP, Kazal, immunoglobulin, Kunitz and NTR domain-containing protein 2	Q8TEU8	WFIKKN2	0.48 (0.41–0.57)	2.52E−18	0.65 (0.56–0.75)	3.52E−09
Brevican core protein	Q96GW7	BCAN	0.50 (0.42–0.59)	4.72E−15	0.61 (0.51–0.73)	1.87E−07
Neural cell adhesion molecule 1, 120 kDa isoform	P13591	NCAM1	0.51 (0.44–0.60)	6.01E−17	0.63 (0.54–0.73)	1.10E−09
Tumor necrosis factor-inducible gene 6 protein	P98066	TNFAIP6	0.51 (0.44–0.60)	2.05E−15	0.65 (0.56–0.75)	5.30E−09
Wnt inhibitory factor 1	Q9Y5W5	WIF1	0.52 (0.43–0.62)	2.39E−12	0.76 (0.66–0.88)	1.62E−04
Hepatocyte growth factor receptor	P08581	MET	0.52 (0.44–0.60)	1.22E−16	0.73 (0.63–0.85)	4.76E−05
Osteomodulin	Q99983	OMD	0.53 (0.45–0.62)	2.38E−14	0.75 (0.65–0.87)	1.21E−04
Transforming growth factor beta receptor type 3	Q03167	TGFBR3	0.53 (0.45–0.62)	9.27E−16	0.72 (0.62–0.84)	1.78E−05
Mast/stem cell growth factor receptor Kit	P10721	KIT	0.56 (0.47–0.65)	9.54E−13	0.74 (0.64–0.86)	5.65E−05
Gelsolin	P06396	GSN	0.57 (0.49–0.67)	1.83E−12	0.66 (0.57–0.76)	3.50E−08
Iduronate 2-sulfatase	P22304	IDS	0.59 (0.51–0.69)	2.21E−11	0.76 (0.65–0.87)	1.70E−04
72 kDa type IV collagenase	P08253	MMP2	0.60 (0.51–0.69)	2.09E−11	0.64 (0.56–0.74)	3.08E−09
Neural cell adhesion molecule L1-like protein	O00533	CHL1	0.60 (0.51–0.69)	2.41E−11	0.73 (0.63–0.83)	6.26E−06
Epidermal growth factor receptor	P00533	EGFR	0.60 (0.51–0.70)	7.17E−11	0.73 (0.62–0.86)	2.28E−04
Interleukin-1 Receptor accessory protein	Q9NPH3	IL1RAP	0.60 (0.51–0.70)	1.61E−10	0.71 (0.62–0.82)	1.61E−06
Apolipoprotein B	P04114	APOB	0.60 (0.52–0.69)	9.93E−12	0.77 (0.67–0.88)	1.66E−04

Table 2 (continued)

Protein full name	UniProt	Gene symbol	KORA (n = 998)		HUNT (n = 923)	
			OR (95% CI)	P-value	OR (95% CI)	P-value
Neurogenic locus notch homolog protein 1	P46531	NOTCH1	0.61 (0.52–0.70)	4.64E−11	0.77 (0.67–0.88)	1.63E−04
Plasma protease C1 inhibitor	P05155	SERPING1	0.61 (0.53–0.71)	3.89E−11	0.66 (0.56–0.77)	1.80E−07
Chordin-like protein 1	Q9BU40	CHRDL1	0.67 (0.57–0.79)	2.86E−06	0.72 (0.60–0.86)	4.50E−04
Thrombin	P00734	F2	0.68 (0.57–0.81)	2.65E−05	0.65 (0.55–0.75)	4.13E−08
Kallikrein-8	O60259	KLK8	0.72 (0.62–0.83)	5.94E−06	0.79 (0.69–0.90)	4.12E−04
Superoxide dismutase [Mn], mitochondrial	P04179	SOD2	0.73 (0.63–0.84)	2.42E−05	0.74 (0.64–0.85)	3.13E−05
Muellerian-inhibiting factor	P03971	AMH	0.74 (0.64–0.85)	4.28E−05	0.77 (0.66–0.88)	4.05E−04

All analyses were adjusted for age, sex, smoking status and physical activity

OR; odds ratio per 1 SD increase in log-transformed protein levels

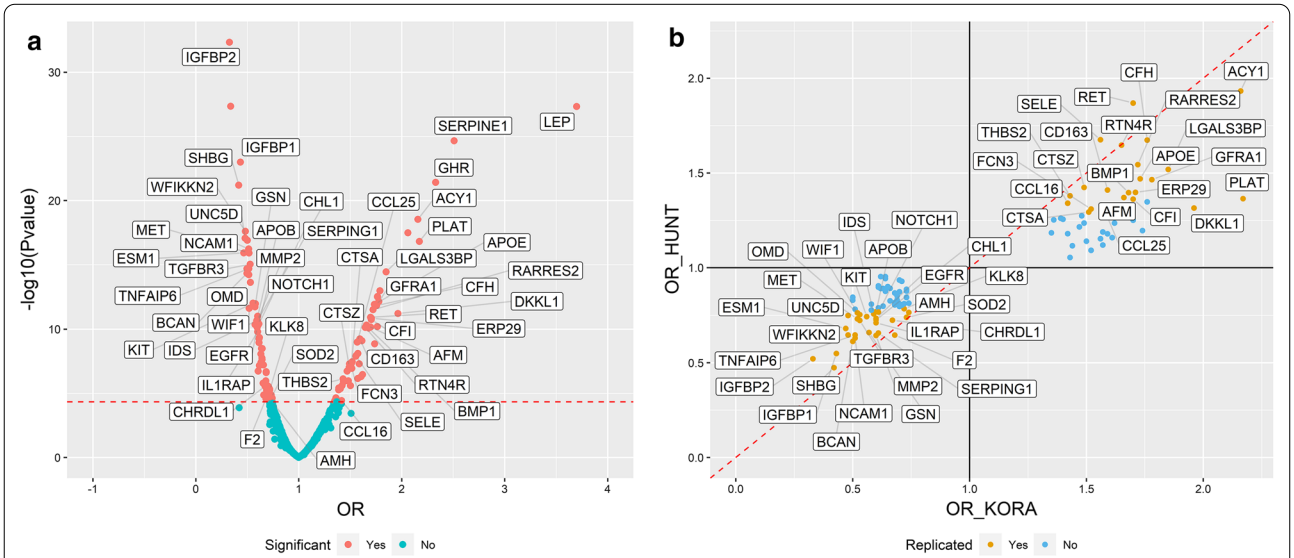

Fig. 1 Results of proteome-wide analysis of prevalent MetS, with replicated proteins labelled by their gene name. **a** Volcano plot of the results in KORA. **b** Concordance plot examining effect sizes in KORA and HUNT. OR; odds ratio per 1 SD increase in log-transformed protein levels

(95% CI = 0.71–0.79). Comparing the LASSO selected predictive model to the age and sex model yielded a delta AUC of 0.12 in KORA, which was significant based on the DeLong test (Additional file 1: Figure S2). The top 2 performing protein were netrin receptor (UNC5D) with AUC = 0.66 (CI = 0.66–0.71) and aminoacylase-1 (ACY1) with AUC = 0.65 (CI = 0.60–0.70) (Additional file :2Table S5).

Our investigation in the utility of prevalent MetS protein associations as biomarkers yielded a 15-protein diagnostic model. The model yielded lower performance in HUNT with an AUC-KORA of 0.87 (95% CI = 0.85–0.89) and an AUC-HUNT of 0.74 (95% CI = 0.71–0.77) (Additional file 1).

Mendelian randomization

We explored if the proteins were causal to MetS (Additional file 2: Table S6). Of the 29 proteins for which we found IVs, 3 showed Bonferroni significant causal effects on MetS (Fig. 4), namely apolipoprotein E3 (APOE3) (Wald-Ratio = − 0.12, Wald-p = 3.63e−13), apolipoprotein B (APOB) (Wald-Ratio = − 0.09, Wald-p = 2.54e−04) and proto-oncogene tyrosine-protein kinase receptor (RET) (Wald-Ratio = 0.10, Wald-p = 5.40e−04).

Table 3 Bonferroni significant results of the proteome-wide analysis with incident MetS in KORA (N = 623), sorted by the magnitude of the OR

Target full name	UniProt	Gene symbol	OR (CI)	P-value
Plasminogen activator inhibitor 1	P05121	SERPINE1	1.82 (1.46–2.30)	2.28E−07
Growth hormone receptor	P10912	GHR	1.65 (1.33–2.04)	4.63E−06
Aminoacylase-1	Q03154	ACY1	1.64 (1.30–2.09)	4.02E−05
C5a anaphylatoxin	P01031	C5	1.62 (1.32–2.01)	6.52E−06
Adiponectin	Q15848	ADIPOQ	0.55 (0.43–0.70)	1.83E−06
Insulin-like growth factor-binding protein 2	P18065	IGFBP2	0.55 (0.44–0.68)	8.36E−08
WAP, Kazal, immunoglobulin, Kunitz and NTR domain-containing protein 2	Q8TEU8	WFIKKN2	0.58 (0.46–0.73)	4.09E−06
Netrin receptor UNC5D	Q6UXZ4	UNC5D	0.61 (0.48–0.77)	2.82E−05
Sex hormone-binding globulin	P04278	SHBG	0.61 (0.49–0.77)	3.43E−05
Iduronate 2-sulfatase	P22304	IDS	0.63 (0.50–0.77)	1.84E−05
Hepatocyte growth factor receptor	P08581	MET	0.63 (0.50–0.77)	1.78E−05
Advanced glycosylation end product-specific receptor, soluble	Q15109	AGER	0.63 (0.51–0.77)	1.10E−05
Insulin-like growth factor-binding protein 1	P08833	IGFBP1	0.63 (0.51–0.79)	4.26E−05
Interleukin-1 receptor type 1	P14778	IL1R1	0.64 (0.52–0.79)	2.64E−05

All analyses were adjusted for age, sex, smoking status and physical activity

OR; odds ratio per 1 SD increase in log-transformed protein levels

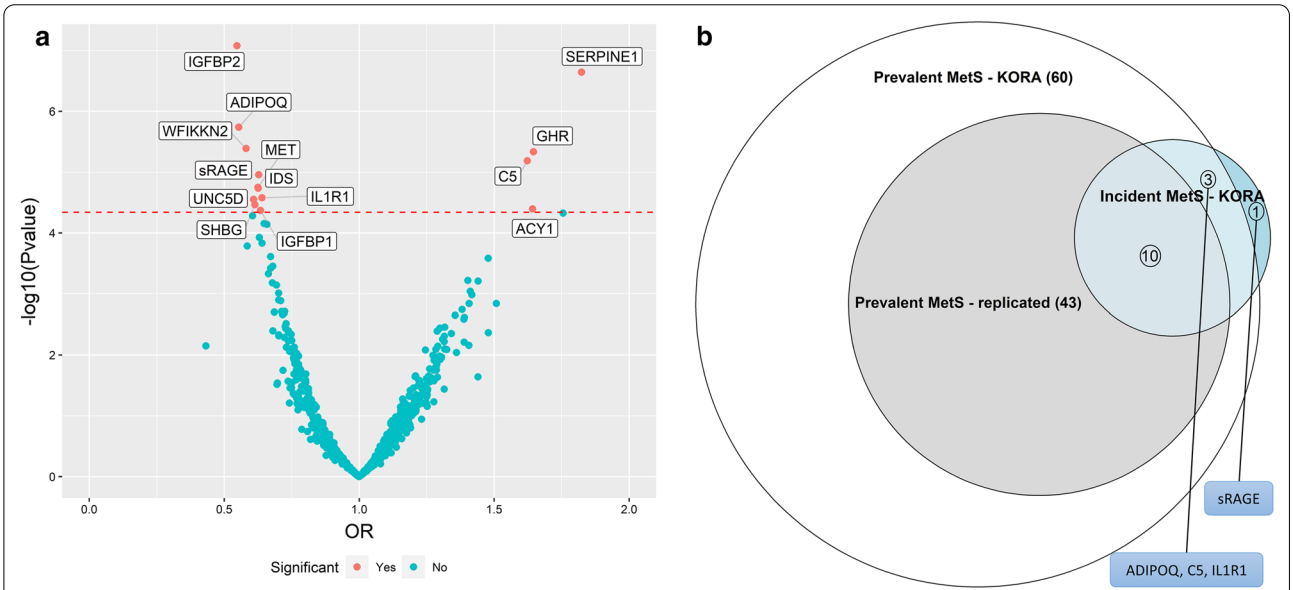

Fig. 2 Results of proteome-wide analysis of incident MetS in KORA. **a** Volcano plot with Bonferroni significant proteins labelled by their gene name. **b** Euler diagram showing extent of overlap between incident MetS results in KORA and prevalent MetS results in KORA and its replicated results in HUNT. OR; odds ratio per 1 SD increase in log-transformed protein levels

Discussion

We used aptamer-based proteomics to investigate plasma protein associations with prevalent and incident MetS and, for those proteins showing a relationship with this syndrome, examined their utility as biomarkers and assessed their causal relationship with MetS. Of the 116 proteins associated with prevalent MetS in the KORA F4 study, 53 successfully replicated in the HUNT3 study. The proteins with the largest effect estimates were leptin and IGFBP2, both of which have been previously reported to be associated with obesity, T2D and MetS [35–38]. The replicated results also included 30 new

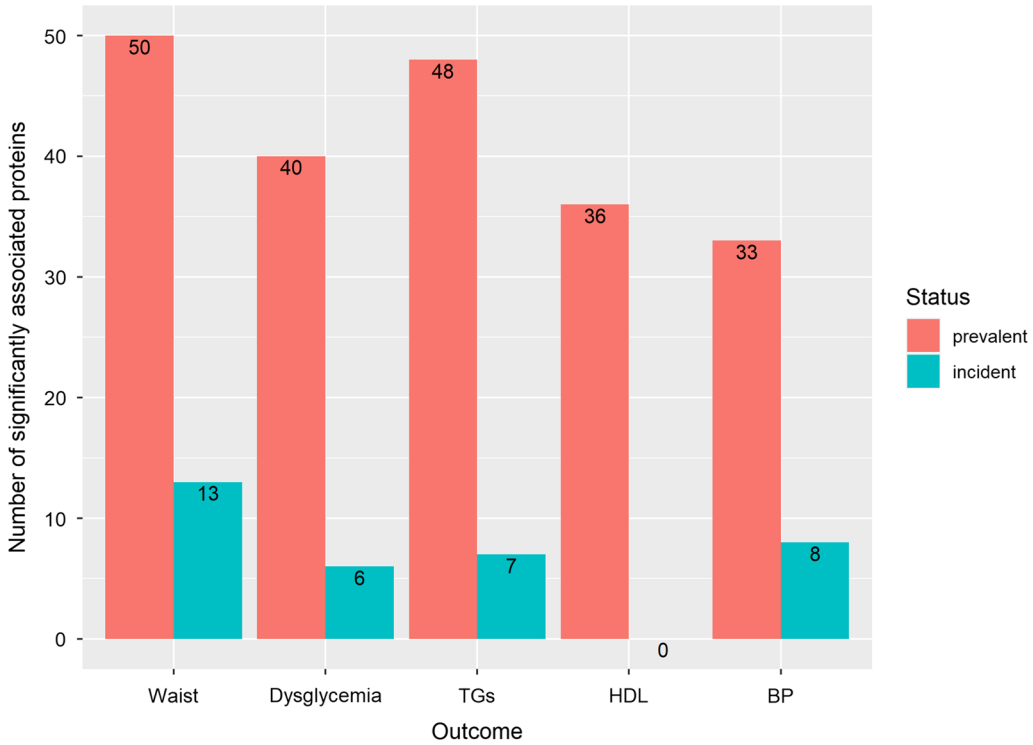

Fig. 3 Barplot showing protein associations of prevalent and incident individual MetS components with replicated prevalent MetS results (n = 53 proteins) and KORA incident MetS results (n = 14 proteins) respectively. The abbreviations used in the y-axis for the MetS components are: Waist (increased waist circumference); dysglycemia (increased blood glucose level); TGs (hypertriglyceridemia); HDL (reduced HDL); BP (increased blood pressure)

protein associations, not previously reported to be associated with MetS, including neural cell adhesion molecule L1-like protein (CHL1), complement factor I (CFI), GDNF family receptor alpha-1 (GFRA1), kallikrein-8 (KLK8), brevican core protein (BCAN), dickkopf-like protein 1 (DKKL1), netrin receptor (UNC5D), NTR domain-containing protein 2 (WFIKKN2), and endoplasmic reticulum protein 29 (ERP29).

Replicated proteins overlap with the protein associations with body mass index (BMI) and type 2 diabetes. WFIKKN2, a protease inhibitor, was reported to be negatively associated with BMI with potential bi-directional causal effect as demonstrated by MR analysis [36]. ERP29, a chaperone protein, has been reported to be positively associated with BMI and to have a role in proinsulin secretion [39]. Of the replicated proteins, endothelial cell-specific molecule 1 (ESM1), has been reported to be low in liver steatosis in MetS patients [40] and in macroalbuminuria in T2D patients [41], both of which are in line with the negative association between ESM1 and MetS observed here. However, ESM1 was reported to be positively associated with atherosclerotic CVD [42].

Associations with incident MetS overlapped with prevalent MetS results, except for sRAGE, which was unique to incident MetS. sRAGE acts as a decoy of the RAGE cell surface receptor. sRAGE exogenously traps advanced glycation end products, therefore decreasing their harmful inflammatory effects through the blockage of their action on RAGE. sRAGE has been previously reported to be inversely associated with T2D, BMI and MetS [36, 43, 44] and was reported to lower the risk of CVD in T2D patients through the modulation of cardiovascular cell apoptosis [45]. RAGE-knockout mice were shown to suffer from accelerated weight gain, hypercholesterolemia and increased insulin levels pointing to the potential complex role of the RAGE family of receptors in the pathogenesis of insulin resistance and obesity [46].

To assess which of the MetS components are driving our observed results, we explored potential associations between our replicated protein associations with prevalent MetS and the respective MetS components and between incident MetS significant proteins and the respective MetS components. In total, 18 of the 53 replicated proteins were associated with all prevalent components. Of them, five were previously reported to be associated with all MetS components namely leptin, IGFBP1, IGFBP2, tissue-type plasminogen activator (PLAT) and SERPINE1 [12]. Ten of these 18 proteins

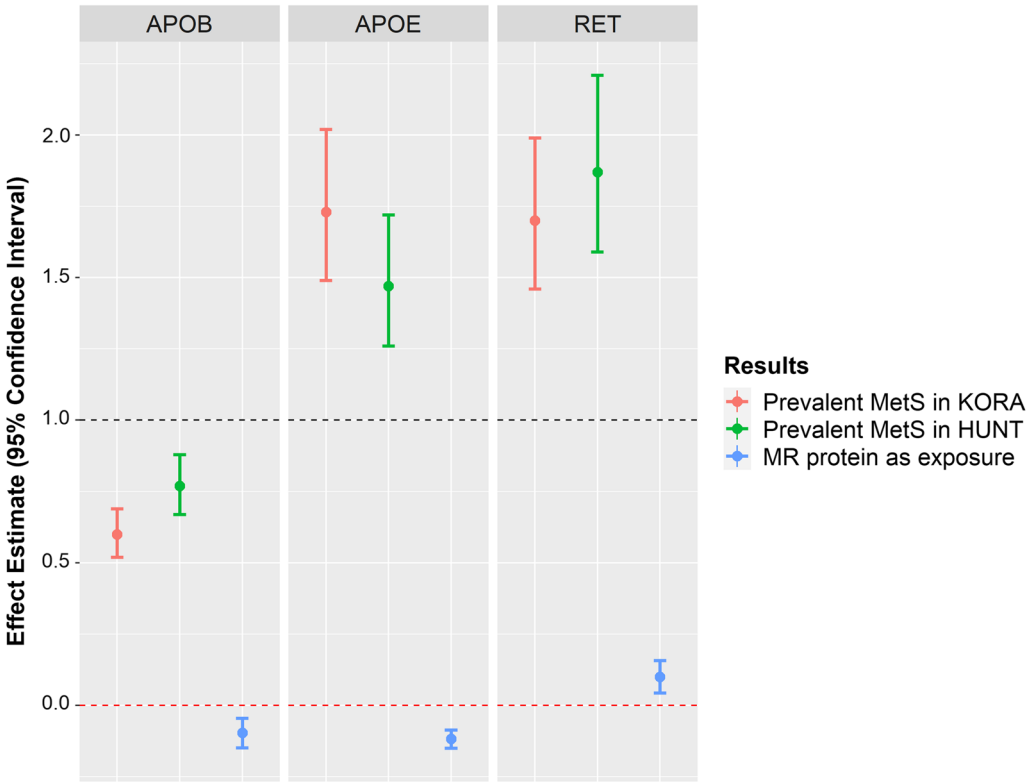

Fig. 4 Mendelian randomization results with MetS as outcome compared with observational effect estimates. Effect estimates represent odds ratios for association results and represent beta coefficients for MR with proteins as exposure

(leptin, IGFBP1, IGFBP2, SHBG, growth hormone receptor (GHR), hepatocyte growth factor receptor (MET), galectin-3-binding protein (LG3BP), APOB and Wnt inhibitory factor 1 (WIF-1)) were previously reported to be associated with T2D, reflecting the shared pathogenic pathways between the two entities [47, 48]. Moreover, 4 of the 18 proteins (PLAT, SERPINE1, 72 kDa type IV collagenase (MMP2), NCAM1) have been reported to be associated with CVD, providing further evidence for the link between CVD and MetS. However, in the present study MMP2 showed a negative association with MetS, contradicting previous reports [49]. While MMP2 has been reported to be increased in metabolic syndrome and cardiovascular disease, its deficiency has also been reported to be associated with metabolic and inflammatory pathologies, pointing toward a complex relationship of MMP2 with cardiometabolic disorders [50, 51].

Of the incident MetS components, ADIPOQ and IGFBP2 were common to all incident components except for reduced HDL, which showed no protein associations. ADIPOQ and IGFBP2 were reported before to be associated with T2D and obesity [36, 52, 53].

Moreover, we evaluated the performance of the proteins as prediction biomarkers, both as a risk score and as individual biomarkers. As predictors of future MetS, the risk score had moderate performance in KORA (AUC = 0.75). As single predictive biomarkers the top five proteins included UNC5D, ACY1, SERPINE1, sRAGE and C5a anaphylatoxin. The lower performance of the proteins as biomarkers could be partly attributed to the differences in baseline characteristics of both cohorts and to the definition of the MetS, which relies on arbitrarily defined cut-off points based on risk assessment of its different components.

The investigation into the causal effects of proteins on MetS showed evidence for 2 protective casual proteins—APOE3 and APOB—and one harmful, RET. Except for APOE3, the causal effect of the proteins had the same effect direction as their observational results. APOE3 is an isoform of the APOE gene, which is a protein-coding gene with two other isoforms, namely APOE2 and APOE4. The APOE isoforms are encoded by two SNPs namely rs429358-C/T and rs7412-C/T. The combination of rs429358-T and rs7412-T is characteristic of the second isoform, of rs429358-T and rs7412-C is characteristic of the third isoform and of rs429358-C and rs7412-C is characteristic of the fourth isoform [54].

The discrepancy between MR and observational results of APOE3 could be due to the fact that the causal effect represents lifetime exposure in comparison to the observational time point effect. Additionally, the IV used in the MR analysis rs1065853 with the effect allele T, is in LD with the T allele of the SNP rs7412. The T allele of the SNP rs7412 characterizes the genotype of the APOE2 polymorphism, indicating that the MR result reflects the effect of APOE2 and not APOE3.APOE2 has been reported to be associated with lower risk of MetS in Uyghur ethnic men [55], with longevity [56] and with lower risk of Alzheimer's disease.

Strengths and limitations

Through the use of the high throughput aptamer based SOMAscan platform, we assessed the association of MetS with a large number of proteins (1095 in total). The use of plasma samples allowed us to find associations which may reflect the processes of multiple tissues and pathways that may be involved in the pathogenesis of MetS; as plasma is easily accessible, our discovered associations may be more readily transferable for use as clinical biomarkers. The replication in the HUNT study indicates broader generalizability of our results. The application of MR to decipher the causal framework governing these associations will enable future investigators to prioritize our results toward drug target identification and further functional investigation of MetS.

There are a number of limitations to our study. We were not able to apply the same MetS definition in the replication study HUNT as in the discovery KORA as the former lacked fasting blood sample collection; however, we used clinically defined cut-off points of non-fasting measurements that reflect the same pathologies identified using fasting measurements. Notably, a study comparing MetS-scores defined using fasting vs. non-fasting samples found that both scores were linked to the development of coronary artery disease and diabetes [57]. The aptamer-based technique could suffer from cross-reactivity; however, our results included proteins replicating previously reported associations measured using techniques other than SOMAscan [58]. The analysis of incident MetS were conducted in a smaller sample size than prevalent MetS and we could not investigate replication in HUNT due to the lack of follow-up data.

We applied rigorous methods in our causality analysis using MR to use only valid IVs and to apply sensitivity analyses to evaluate pleiotropy. However, MR is dependent on multiple assumptions that are hard to verify and test and its results should be interpreted with caution. Moreover, the studies we used in the causal analyses differed in power for the exposures and the outcome, with MetS GWAS having a bigger sample size and subsequently more power than the protein GWAS studies.

Conclusion

Our results provide a comprehensive analysis of the associations between plasma proteins and MetS. Replicated results included proteins previously reported to be associated with cardio-metabolic traits, thus pointing to pathogenic pathways they share with MetS, including insulin resistance and CVD. These proteins include leptin, GHR, SHBG, IGFBP1 and IGFBP2. Replicated results also included proteins involved in the pathogenesis of CVD, such as PLAT, SERPINE1 and members of the complement system. Our replicated results identified new proteins including ERP29, KLK8, DKKL1 and WFIKKN2. We identify sRAGE to be uniquely associated with the incidence of MetS, which is in line with the observed phenotype in sRAGE knockout mice models.

Biomarker analysis identified an eight proteins predictive panel with an AUC of 0.75. Moreover, causal analysis using Mendelian randomization suggested causal effects of APOE2, APOB and RET on MetS. Further functional studies are needed to clarify their roles in the pathogenesis of MetS.

Abbreviations

MetS: Metabolic syndrome; T2D: Type 2 diabetes; CVD: Cardiovascular diseases; CAD: Coronary artery disease; MR: Mendelian randomization; KORA: Cooperative health research in the Region of Augsburg; HUNT: The Nord-Trøndelag Health Study; HDL: High density lipoprotein cholesterol; LASSO: Least absolute shrinkage and selection operator; ROC-AUC: Area under the receiver operating characteristic curve; SNP: Single nucleotide polymorphism; GWAS: Genome wide association studies; BMI: Body mass index.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12933-021-01299-2.

Additional file 1: Figure S1. Pearson's correlation plot of replicated proteins in: A) KORA; B) HUNT. **Figure S2.** ROC curve comparing the bootstrap ranking LASSO selected protein model with age and sex model predicting incident MetS in KORA, showing the AUCs, their 95% CI and the difference (delta AUC) and p-value of the DeLong test comparing both models. **Figure S3.** ROC curve comparing the bootstrap ranking LASSO selected protein model with age and sex model predicting prevalent MetS in KORA (A) and HUNT (B), showing the AUCs, their 95% CI and the difference (delta AUC) and p-value of the DeLong test comparing both models. **Figure S4.** Calibration plots of the bootstrap ranking LASSO-selected MetS diagnostic model in: A) KORA; B) HUNT. **Figure S5.** STRING protein-protein interaction network constructed using the prevalent or incident MetS associated proteins in KORA without adding additional interactor proteins.

Additional file 2: Table S1. Proteins significantly associated with prevalent MetS in KORA, their corresponding results in HUNT as well as their random effect meta-analysis results. **Table S2.** Overlap between the results of prevalent MetS components and replicated prevalent MetS results. **Table S3.** Overlap between the results of incident MetS components and incident MetS KORA results. **Table S4.** Coefficients of

the incident MetS predictive protein risk score. **Table S5.** Results of the 14 significantly associated proteins with incident MetS as single predictive biomarkers in KORA. **Table S6.** Results of Mendelian randomization analysis with proteins as the exposure and MetS as the outcome. **Table S7.** Coefficients of the prevalent MetS diagnostic protein risk score. **Table S8.** Results of the 116 significantly associated proteins with prevalent MetS as single diagnostic biomarkers in KORA and HUNT.

Acknowledgements
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. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ.

Authors' contributions
MAE participated in the design of the study, analyzed the data, interpreted the results, and wrote and revised the manuscript. MW conceived the research question, participated in its design and contributed to interpretation of the results. AP is responsible for the overall design of the KORA cohort, participated in the design of the study and contributed to interpretation of the results. SZ, RW and CG helped with the analyses. AP, MW, RW, SZ, CH, CG, JG, WR, WK, MFS, KH, KS, BT and CJ were involved in the data collection, data management, and preparation of their respective cohorts. MAE, MW and AP 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. All authors read and approved the final manuscript.

Funding
Open Access funding enabled and organized by Projekt DEAL. MW was supported 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). KS and SZ were supported by the Biomedical Research Program at Weill Cornell Medicine in Qatar, a program funded by the Qatar Foundation, also by Qatar National Research Fund (QNRF) Grant NPRP11C-0115-180010.

Availability of data and materials
The KORA data is available through application at the KORA Project Application Self-Service Tool (https://epi.helmholtz-muenchen.de/). The HUNT data is available through application to the HUNT Research Centre (http://www.ntnu.edu/hunt/data). Data used in the two-sample MR analysis are publicly available and can be accessed through: MetS GWAS by Lind using the GWAS Catalog accession (GCST009602); Sun et al. at: http://www.phpc.cam.ac.uk/ceu/proteins/; Suhre et al. at: http://proteomics.gwas.eu; and Emilsson et al. at www.sciencemag.org/content/361/6404/769/suppl/DC1.

Declarations

Ethics approval and consent to participate
The KORA study was approved by the ethics committee of the Bavarian Medical Association. The HUNT study was approved by the Regional Committee for Medical and Health Research Ethics. In both studies, written informed consent was obtained from each participant.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Research Unit 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, Munich, Germany. [4]Weill Cornell Medicine-Qatar, Education City, PO Box 24144, Doha, Qatar. [5]German Center for Diabetes Research (DZD), München-Neuherberg, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany. [6]Biomolecular Mass Spectrometry, Max Planck Institute for Heart and Lung Research, Ludwigstrasse 43, 61231 Bad Nauheim, Germany. [7]The German Centre for Cardiovascular Research (DZHK), Partner Site Rhine-Main, Max Planck Institute for Heart and Lung Research, Bad Nauheim, Germany. [8]Institute of Biometrics and Epidemiology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany. [9]Deutsches Herzzentrum München, Technische Universität München, Munich, Germany. [10]Institute of Epidemiology and Medical Biometry, University of Ulm, Ulm, Germany. [11]Department of Medicine I, University Hospital Munich, Ludwig-Maximilians-University, Munich, Germany. [12]K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, NTNU - Norwegian University of Science and Technology, Trondheim, Norway. [13]HUNT Research Center, Department of Public Health, NTNU - Norwegian University of Science and Technology, Levanger, Norway. [14]Chair of Epidemiology, Institute for Medical Information Processing, Biometry and Epidemiology, Medical Faculty, Ludwig-Maximilians-Universität München, Munich, Germany.

Received: 16 March 2021 Accepted: 5 May 2021
Published online: 20 May 2021

References

1. Rask-Madsen C, Kahn CR. Tissue-specific insulin signaling, metabolic syndrome, and cardiovascular disease. Arterioscler Thromb Vasc Biol. 2012;32(9):2052–9.
2. Tune JD, Goodwill AG, Sassoon DJ, Mather KJ. Cardiovascular consequences of metabolic syndrome. Transl Res J Lab Clin Med. 2017;183:57–70.
3. Afshin A, Forouzanfar MH, Reitsma MB, Sur P, Estep K, Lee A, et al. Health effects of overweight and obesity in 195 countries over 25 years. N Engl J Med. 2017;377(1):13–27.
4. O'Neill S, O'Driscoll L. Metabolic syndrome: a closer look at the growing epidemic and its associated pathologies. Obes Rev Off J Int Assoc Study Obes. 2015;16(1):1–12.
5. International Diabetes Federation: The IDF consensus worldwide definition of the metabolic syndrome. http://www.idf.org/metabolic-syndrome.
6. Hanefeld M, Schaper F, Ceriello A. Geschichte und Definition(en) des metabolischen Syndroms. Der Internist. 2007;48(2):117–25.
7. Grundy SM. Metabolic syndrome: a multiplex cardiovascular risk factor. J Clin Endocrinol Metab. 2007;92(2):399–404.
8. Kahn R. Metabolic syndrome is it a syndrome? Does it matter? Circulation. 2007;115(13):1806–11.
9. Richard CC. The metabolic syndrome: is it really a syndrome? Clin Cardiol. 2006;29(12):523–4.
10. Abifadel M, Varret M, Rabès JP, Allard D, Ouguerram K, Devillers M, et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nat Genet. 2003;34(2):154–6.
11. Lind L. Genome-wide association study of the metabolic syndrome in UK biobank. Metab Syndr Relat Disord. 2019;17(10):505–11.
12. Lind L, Elmståhl S, Ingelsson E. Cardiometabolic proteins associated with metabolic syndrome. Metab Syndr Relat Disord. 2019;17(5):272–9.
13. Priest C, Tontonoz P. Inter-organ cross-talk in metabolic syndrome. Nat Metab. 2019;1(12):1177–88.
14. Robberecht H, Hermans N. Biomarkers of metabolic syndrome: biochemical background and clinical significance. Metab Syndr Relat Disord. 2016;14(2):47–93.
15. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018. https://doi.org/10.7554/eLife.34408.
16. Holle, R.; Happich, M.; Lowel, H.; Wichmann, H.E. KORA--a research platform for population based health research. *Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes (Germany)) 2005, 67 Suppl 1*, S19-25. https://doi.org/10.1055/s-2005-858235

17. Krokstad S, Langhammer A, Hveem K, Holmen TL, Midthjell K, Stene TR, et al. Cohort profile: the HUNT study, Norway. Int J Epidemiol. 2013;42(4):968–77.
18. Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. PLoS ONE. 2010;5(12):e15004.
19. Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. Nat Commun. 2017;8:14357.
20. Ganz P, Heidecker B, Hveem K, et al. Development and validation of a protein-based risk score for cardiovascular outcomes among patients with stable coronary heart disease. JAMA. 2016;315(23):2532–41.
21. Alberti KG, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. Circulation. 2009;120(16):1640–5.
22. Driver SL, Martin SS, Gluckman TJ, Clary JM, Blumenthal RS, Stone NJ. Fasting or nonfasting lipid measurements: it depends on the question. J Am Coll Cardiol. 2016;67(10):1227–34.
23. American DA. Diagnosis and classification of diabetes mellitus. Diabetes Care. 2010;33(Suppl 1):S62–9.
24. Guo P, Zeng F, Hu X, Zhang D, Zhu S, Deng Y, et al. Improved variable selection algorithm using a LASSO-type penalty, with an application to assessing hepatitis B infection relevant factors in community residents. PLoS ONE. 2015;10(7):e0134151.
25. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformat. 2011;12(1):77.
26. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21(1):128–38.
27. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837–45.
28. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607–13.
29. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature. 2018;558(7708):73–9.
30. Emilsson V, Ilkov M, Lamb JR, Finkel N, Gudmundsson EF, Pitts R, et al. Co-regulatory networks of human serum proteins link genetics to disease. Science. 2018;361(6404):769–73.
31. Hartwig FP, Davies NM, Hemani G, Davey SG. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. Int J Epidemiol. 2017;45(6):1717–26.
32. Teumer A. Common methods for performing Mendelian randomization. Front Cardiovasc Med. 2018;5:51.
33. Zheng J, Baird D, Borges MC, Bowden J, Hemani G, Haycock P, et al. Recent developments in Mendelian randomization studies. Curr Epidemiol Rep. 2017;4(4):330–45.
34. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44(2):512–25.
35. Katsiki N, Mikhailidis DP, Banach M. Leptin, cardiovascular diseases and type 2 diabetes mellitus. Acta Pharmacol Sin. 2018;39(7):1176–88.
36. Zaghlool SB, Sharma S, Molnar M, Matías-García PR, Elhadad MA, Waldenberger M, et al. Revealing the role of the human blood plasma proteome in obesity using genetic drivers. Nat Commun. 2021;12(1):1279.
37. Wittenbecher C, Ouni M, Kuxhaus O, Jähnert M, Gottmann P, Teichmann A, et al. Insulin-like growth factor binding protein 2 (IGFBP-2) and the risk of developing type 2 diabetes. Diabetes. 2019;68(1):188–97.
38. Pouriamehr S, Barmaki H, Rastegary M, Lotfi F, Nabi AM. Investigation of insulin-like growth factors/insulin-like growth factor binding proteins regulation in metabolic syndrome patients. BMC Res Notes. 2019;12(1):653.
39. Viviano J, Brecker M, Ferrara-Cook C, Suaud L, Rubenstein RC. ERp29 as a regulator of insulin biosynthesis. PLoS ONE. 2020;15(5):e0233502.
40. Erman H, Beydogan E, Cetin SI, Boyuk B. Endocan: a biomarker for hepatosteatosis in patients with metabolic syndrome. Mediators Inflamm. 2020;2020:3534042.
41. Cikrikcioglu MA, Erturk Z, Kilic E, Celik K, Ekinci I, Yasin Cetin AI, et al. Endocan and albuminuria in type 2 diabetes mellitus. Ren Fail. 2016;38(10):1647–53.
42. Balta S, Mikhailidis DP, Demirkol S, Ozturk C, Celik T, Iyisoy A. Endocan: a novel inflammatory indicator in cardiovascular disease? Atherosclerosis. 2015;243(1):339–43.
43. He CT, Lee CH, Hsieh CH, Hsiao FC, Kuo P, Chu NF, et al. Soluble form of receptor for advanced glycation end products is associated with obesity and metabolic syndrome in adolescents. Int J Endocrinol. 2014;2014:657607.
44. Miranda ER, Somal VS, Mey JT, Blackburn BK, Wang E, Farabi S, et al. Circulating soluble RAGE isoforms are attenuated in obese, impaired-glucose-tolerant individuals and are associated with the development of type 2 diabetes. Am J Physiol Endocrinol Metab. 2017;313(6):E631–40.
45. Tsoporis JN, Hatziagelaki E, Gupta S, Izhar S, Salpeas V, Tsiavou A, et al. Circulating ligands of the receptor for advanced glycation end products and the soluble form of the receptor modulate cardiovascular cell apoptosis in diabetes. Molecules. 2020;25(22):5235.
46. Leuner B, Max M, Thamm K, Kausler C, Yakobus Y, Bierhaus A, et al. RAGE influences obesity in mice. Effects of the presence of RAGE on weight gain, AGE accumulation, and insulin levels in mice on a high fat diet. Z Gerontol Geriatr. 2012;45(2):102–8.
47. Alessi MC, Juhan-Vague I. PAI-1 and the metabolic syndrome: links, causes, and consequences. Arterioscler Thromb Vasc Biol. 2006;26(10):2200–7.
48. Choi H, Koh HWL, Zhou L, Cheng H, Loh TP, Parvaresh Rizi E, Toh SA, Ronnett GV, Huang BE, Khoo CM. Plasma protein and microRNA biomarkers of insulin resistance: a network-based integrative -omics analysis. Front Physiol. 2019. https://doi.org/10.3389/fphys.2019.00379.
49. Erman H, Gelisgen R, Cengiz M, Tabak O, Erdenen F, Uzun H. The association of vascular endothelial growth factor, metalloproteinases and their tissue inhibitors with cardiovascular risk factors in the metabolic syndrome. Eur Rev Med Pharmacol Sci. 2016;20(6):1015–22.
50. Fernandez-Patron C, Kassiri Z, Leung D. Modulation of systemic metabolism by MMP-2: from MMP-2 deficiency in mice to MMP-2 deficiency in patients. Compr Physiol. 2016;6(4):1935–49.
51. Hardy E, Hardy-Sosa A, Fernandez-Patron C. MMP-2: is too low as bad as too high in the cardiovascular system? Am J Physiol Heart Circ Physiol. 2018;315(5):H1332–40.
52. Elhadad MA, Jonasson C, Huth C, Wilson R, Gieger C, Matias P, et al. Deciphering the plasma proteome of type 2 diabetes. Diabetes. 2020;141:db200296.
53. Gudmundsdottir V, Zaghlool SB, Emilsson V, Aspelund T, Ilkov M, Gudmundsson EF, et al. Circulating protein signatures and causal candidates for type 2 diabetes. Diabetes. 2020;69(8):1843–53.
54. Abondio P, Sazzini M, Garagnani P, Boattini A, Monti D, Franceschi C, et al. The genetic variability of APOE in different human populations and its implications for longevity. Genes (Basel). 2019;10(3):222.
55. Sun Y, Wei R, Yan D, Xu F, Zhang X, Zhang B, et al. Association between APOE polymorphism and metabolic syndrome in Uyghur ethnic men. BMJ Open. 2016;6(1):e010049.
56. Shinohara M, Kanekiyo T, Tachibana M, Kurti A, Shinohara M, Fu Y, et al. APOE2 is associated with longevity independent of Alzheimer's disease. Elife. 2020;9:62199.
57. DeBoer MD, Filipp SL, Gurka MJ. Associations of a metabolic syndrome severity score with coronary heart disease and diabetes in fasting vs. non-fasting individuals. Nutr Metab Cardiovasc Dis NMCD. 2020;30(1):92–8.
58. Suhre K, McCarthy MI, Schwenk JM. Genetics meets proteomics: perspectives for large population-based studies. Nat Rev Genet. 2021;22(1):19–37.

Publisher's Note

Acknowledgements

At the end of my thesis, I would like to acknowledge and thank my supervisors, team members, friends, colleagues, and family for their continuous support and help.

First, I would like to thank my supervisor Prof. Annette Peters for her advices and insightful guidance that she has been providing me throughout. A special thank you to Dr. Melanie Waldenberger, who saw a potential in me that I did not know exist, helped me see and achieve it. I am grateful for her help and support that started with her supervision of my master thesis and continued thereafter.

I am thankful to the members of my thesis committee, Prof. Dr. med. Wolfgang Koenig und PD Dr. med. Moritz Sinner, who provided me with much needed clinical and scientific input to my research as well as to my career.

I am indebted to the team of the research unit of molecular epidemiology (AME), my team members and friends, Rory Wilson, Pamela Matias, Nadine Lindemann, Dr. Christian Gieger and Dr. Harald Grallert and the rest of the team for their input, scientific support and for the enlightening discussions we exchanged.

I am grateful to the team of the institute of epidemiology in Helmholtz center in Munich, Dr. Cornelia Huth and Prof. Barbara Thorand, for their scientific support and help, as experts in the field, in my research with type 2 diabetes and metabolic syndrome.

I am indebted to Dr. Gabi Kastenmüller, Prof. Karsten Suhre, Dr. Shaza Zaghlool and Dr. Johannes Graumann, who supported my efforts in analyzing new levels of molecular epidemiology and provided me with much needed expertise in proteomics and molecaular epidemiology.

I thank the co-authors and collaborators from QMDiab and HUNT for their cooperation and initiative throughout our collaborations.

I would like to express my deepest gratitude to the study participants, personnel and funders for their dedication and contribution to science that made work such as this thesis possible.

My university professors and schoolteachers especially Mrs. Amal Allam, my primary school teacher, who supported and guided my early steps in education, encouraged my curiosity and allowed it to grow.

At the end, I would like to thank my family for their support; my mother for her belief in me and her constant encouragement of my curiosity and love of knowledge; my dad for his support; my brothers Mahmoud and Moaz for their companionship and their support even in the worst of days; my wife and soulmate Yasmine for always being there to support and encourage me and my daughter Maha for her smiles that charges my energy and motivates me to keep moving forward.