Proteome-wide assessment of diabetes mellitus in Qatari identifies IGFBP-2 as a risk factor already with early glycaemic disturbances

Noordam, R.; Heemst, D. van; Suhre, K.; Krumsiek, J.; Mook-Kanamori, D.O.

Citation

Noordam, R., Heemst, D. van, Suhre, K., Krumsiek, J., & Mook-Kanamori, D. O. (2020). Proteome-wide assessment of diabetes mellitus in Qatari identifies IGFBP-2 as a risk factor already with early glycaemic disturbances. *Archives Of Biochemistry And Biophysics*, *689*. doi:10.1016/j.abb.2020.108476

Version: Publisher's Version

License: Creative Commons CC BY 4.0 license

Downloaded from: https://hdl.handle.net/1887/3185186

Note: To cite this publication please use the final published version (if applicable).

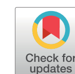
Archives of Biochemistry and Biophysics 689 (2020) 108476

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/locate/yabbi

Proteome-wide assessment of diabetes mellitus in Qatari identifies IGFBP-2 as a risk factor already with early glycaemic disturbances



Raymond Noordam[a,b,*], Diana van Heemst[a], Karsten Suhre[c], Jan Krumsiek[b,d], Dennis O. Mook-Kanamori[e,f]

[a] Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands
[b] Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, Germany
[c] Department of Physiology and Biophysics, Weill Cornell Medicine-Qatar, Doha, Qatar
[d] Department of Physiology and Biophysics, Weill Cornell Medical College, New York, USA
[e] Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands
[f] Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands

ARTICLE INFO

Keywords:
Proteomics
Metabolomics
Diabetes mellitus
Case-control study
Mendelian randomization

ABSTRACT

Background: Proteomics is expected to provide novel insights in the underlying pathophysiology of type 2 diabetes mellitus. In the present study, we aimed to identify and biochemically characterize proteins associated with diabetes mellitus in a Qatari population.

Methods: In a diabetes case-control study (175 cases, 164 controls; Arab, South Asian and Philippine ethnicities), we conducted a discovery study to screen 1141 blood protein levels for associations with diabetes mellitus. Additional analyses were done in controls in relation to Hb1Ac, and biochemical characterization of the main findings was performed with metabolomics (501 metabolites). We performed two-sample Mendelian Randomization to provide evidence of potential causality using data from European descent of the DIAGRAM consortium (74,124 cases of diabetes mellitus and 824,006 controls) for the identified proteins for T2D and Hb1Ac.

Results: After accounting for multiple testing, 30 protein levels were different (p-values $< 8.6e^{-5}$) between cases and controls. Of these, a higher Hb1Ac in controls was associated with a lower IGFBP-2 level (p-value = $4.1e^{-6}$). IGFBP-2 protein level was found lower among cases compared with controls across all ethnicities. In controls, IGFBP-2 was associated with 21 metabolite levels, but specifically connected to the metabolite citrulline in network analyses. We observed no evidence, however, that the association between IGFBP-2 and diabetes mellitus was causal.

Conclusions: We specifically identified IGFBP-2 to be associated with diabetes mellitus, although with no evidence for causality, which was specifically connected to citrulline metabolism.

1. Introduction

The prevalence of type 2 diabetes mellitus (T2D) has increased substantially during the last few decades, and has become a major cause of morbidity and mortality [1]. Many of the cases of T2D can be prevented with changes in lifestyle, such as increased physical activity and compliance to a healthier diet [1–3]. Therefore, much of the current research efforts have been devoted to identify biological markers for early detection as well as biological targets for treatment and prevention [1,4–6].

With efforts from the DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) consortium, in a current sample of approximately 900,000 individuals of European descent, over 240 genetic risk loci have been identified to increase the risk of T2D, explaining 18% of the genetic heritability enhancing the potential for clinical translation [6]. However, direct clinical translation is complicated without knowledge of the translated protein concentrations in relation to T2D. Previously, proteomics analyses have been shown to be an effective way of linking genetic information to disease outcomes [5]. In early proteomics studies, tissue biopsies were used, and provided valuable novel insights of, for example, the role of mitochondrial proteins in muscle tissues in relation to insulin resistance [7]. Due to technological advances, high

* Corresponding author. Department of Internal Medicine, section Gerontology and Geriatrics, Leiden University Medical Center, PO Box 9600, 2300, RC, Leiden, the Netherlands.

E-mail address: r.noordam@lumc.nl (R. Noordam).

https://doi.org/10.1016/j.abb.2020.108476
Received 9 March 2020; Received in revised form 6 June 2020; Accepted 17 June 2020
Available online 22 June 2020
0003-9861/ © 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/BY/4.0/).

throughput proteomics data has become increasingly available in cohort settings [8], and can be used for the identification of novel biomarkers for disease as well as provide further insights in the pathophysiological mechanisms of T2D [9,10].

We hypothesize that a proteomics assessment of T2D will provide us with additional insights that will contribute to our understanding of the pathophysiological mechanisms in T2D. Within the present study, we conducted a protein-wide association study on T2D in a Qatari population, without a specific predefined hypothesis. On the basis of the proteins identified in T2D, we performed further characterization using metabolomics to identify involved biochemical pathways and used a Mendelian Randomization approach to provide evidence of potential causality [11] using instrumental variables derived from published genome-wide association studies on proteomics [12].

2. Methods

2.1. Study setting and design

The present study was embedded in the Qatar Metabolomics Study on Diabetes, a cross-sectional case-control study in which a total of 374 participants were enrolled. The study was realized as a collaboration between the Dermatology Department of the Hamad Medical Corporation (HMC) and the Weill Cornell Medical Collage-Qatar. The design of the study was approved by the Institutional Review Boards of the HMC and Weill Cornell Medical College-Qatar (Research Protocol Number 11131/11). Written informed consent was obtained from all study participants.

Study participants were enrolled for study participation between February and June 2012 at the Dermatology Department of HMC in Doha, Qatar. Inclusion criteria were a primary form of T2D (cases) and an absence of major systemic disorders (controls). Data on five included participants was excluded due to incomplete records, leaving a total original sample of 176 cases and 193 controls. Of the initial 193 controls, 12 had a glycosylated hemoglobin (HbA1c) greater than 6.5% and were classified as cases, which gives a final sample of 188 T2D cases and 181 controls. No matching was performed based on age, sex and ethnicity as it has been previously shown that matching gives similar results as after statistical adjustment for covariates but significantly reduced the statistical power [13,14].

2.2. Phenotyping

Information on age, sex, and T2D history were obtained through questionnaires at the study center. Using standardized protocols, trained researchers determined lightly clothed weight to the nearest decimal with an electronic scale (SECA Scale 813) and height without shoes to the nearest decimal using a stadiometer (SECA Mobile Stadiometer 217). Body mass index (BMI; in kg/m^2) was calculated by dividing weight (in kg) by squared height (in meters). Blood samples were generally taken in non-fasting state between 1:00 and 3:00 p.m. HbA1c was determined at the Department of Laboratory Medicine and Pathology of HMC (Cobas 6000; Roche Diagnostics). Estimated Glomerular filtration rate (eGFR) was calculated using the Cockcroft-Gault formula [15].

2.3. Proteomics measurements

The SOMAscan platform [8] was used to quantify protein levels in non-fasting plasma samples at the WCM-Q proteomics core. No samples were excluded. Protocols and instrumentation were provided and certified using reference samples by SomaLogic Inc. Experiments were conducted under supervision of SomaLogic personnel. Briefly, undepleted EDTA-plasma was diluted into three dilution bins (0.05%, 1%, and 40%) and incubated with bin-specific collections of bead-coupled SOMAmers in a 96-well plate format. Subsequent to washing steps, bead-bound proteins are biotinylated and complexes comprising biotinylated target proteins and fluorescence-labelled SOMAmers are photocleaved off the bead support and pooled. Following recapture on streptavidin beads and further washing steps, SOMAmers are eluted and quantified as a proxy to protein concentration by hybridization to custom arrays of SOMAmer-complementary oligonucleotides. Based on standard samples included on each plate, the resulting raw intensities are processed using a data analysis work flow including hybridization normalization, median signal normalization and signal calibration to control for inter-plate differences. Using this platform, a total of 1141 proteins were measured.

2.4. Metabolomics measurements

Metabolite level quantification was performed in non-fasting plasma samples from all study participants. Within 6 h after sample collection, all samples were centrifuged at 2500 g for 10 min, aliquoted, and stored at −80 °C. Metabolic profiling was achieved using ultrahigh-performance liquid-phase chromatography and gas chromatography separation, coupled with tandem mass spectrometry at Metabolon, Inc. using established procedures [16]. Metabolite levels were scaled by run-day medians, log transformed, and subsequently z-scored (mean = 0, standard deviation = 1).

2.5. Statistical analyses

Characteristics of the study population were assessed separately for T2D cases and controls, and presented as means (with standard deviation), number (with percentage), or median (with interquartile range; for skewed population characteristics only).

Prior to the statistical comparisons between T2D cases and controls, all protein levels were ln2-transformed to approximate a normal distribution. This transformation was found to be most suitable to approximate a normal distribution for most protein levels investigated in the present study.

Primary comparisons between T2D cases and controls were conducted using multivariable linear regression analyses adjusted for age, sex, and body mass index. We used the R-based GeneNet package to provide insights in the protein networks and of the networks of the identified proteins [17], which is based on the partial correlations between the proteins. To prevent selection bias based on T2D status, we conducted the network analyses in controls only. As identified differences in protein levels between T2D cases and controls could be either a cause as well as a consequence of the disease, we repeated the analyses on Hb1Ac in controls only. We assumed that identified proteins present in both analyses are more likely to be in the pathophysiological pathway. Robustness of the associations with T2D of these proteins was further studied across the different ethnicities present in our study population, and with additional adjustment for eGFR. Furthermore, for these proteins, we calculated the partial correlations on T2D.

Of the protein(s) identified with our efforts, we studied the associated metabolic pathways using the metabolomics data. Analyses were performed using linear regression analyses, adjusted for age, sex and body mass index in controls only to prevent biased results. GeneNet was used to provide insights in the networks of metabolites directly connected to the protein being explored [17].

Protein and metabolite levels are frequently interrelated. Conventional correction for multiple testing (e.g., Bonferroni) is therefore likely to be too stringent. Based on the data of the present study, we calculated the number of independent protein/metabolite levels being tested, based on methodology that has been previously described by Li et al. [18], which takes into account the correlation between the protein/metabolite levels. Considering the number of independent levels, we corrected the results for multiple testing with Bonferroni based on the number of independent tests.

All statistical analyses were performed using the R statistical

package (version 3.6.1., www.r-project.org/) and with use of the R-based GeneNet and ggplot2 packages [17,19].

2.6. Mendelian Randomization analyses on type 2 diabetes mellitus

In order to examine whether the most promising proteomics findings from our observational findings are causal risk factors, we conducted two-sample Mendelian Randomization, as previously described [20,21]. As exposures, we selected genetic instruments for the proteins from publicly available summary statistics data of a GWAS done in 3301 individuals of European descent [12]. As genetic instruments, we selected all independent ($R^2 < 0.001$) variants with a p-value $< 1e^{-5}$. As outcome, we used publicly available summary statistics data from a European-ancestry meta-analysis of 32 cohorts including 74,124 cases of T2D and 824,006 controls [6].

For the statistical analyses, we used the well-established methods for two-sample MR analyses; namely inverse-variance weighted (IVW), MR-Egger and median-weighted analyses (MWA), as previously described [20]. The IVW method assumes all genetic instruments are valid and only associate with the study outcome through the exposure. These latter two methods for two-sample MR analyses were specifically developed to test and/or take into account potential bias due to the presence of directional pleiotropy by invalid instrumental variables, and especially MR Egger has a significant reduction in statistical power [22,23]. Analyses were conducted using the TwoSampleMR package implemented in R statistical software [24].

3. Results

3.1. Characteristics of the T2D cases and controls

After exclusion of individuals with missing data on protein levels, the total study sample comprised 175 T2D cases and 164 controls (Table 1). Compared with the control population, T2D cases were on average older (53.1 [SD = 9.8] versus 40.6 [SD = 12.2] years), comprised more women (56 versus 45%) and had on average a higher BMI (30.6 [SD = 6.5] versus 27.6 [SD = 5.4] years).

3.2. Comparison in protein levels between T2D cases and controls

Out of the 1141 protein levels measured in our study population (protein network of the 750 strongest partial correlations displayed in Fig. 1A/B), 584 levels were uncorrelated making us to use a multiple testing-corrected alpha of $8.6e^{-5}$. After accounting for multiple testing and adjustment for sex, age and body mass index, 30 proteins levels were found to be different between T2D and controls (Fig. 1C). Of these, 12 protein levels (e.g., MIA, Adiponectin, Apo-B, and IGFBP-2/6) were lower in T2D cases compared to controls and 18 protein levels (e.g., sE-Selectin, C2/7, and IL-19) were higher in T2D cases compared to controls. The summary statistics of all proteins in relation to T2D can be found in Supplementary Table 1. Based on the 30 identified proteins and the 20 highest partial correlations (range: 0.19–0.47), 22 proteins could be mapped to one out of the five identified protein subnetworks in controls (Fig. 1D).

Table 1
Characteristics of cases with diabetes mellitus and controls in the Qatar Metabolomics Study on Diabetes.

	T2D cases (N = 175)	Controls (N = 164)
Age in years, mean (SD)	53.1 (9.8)	40.6 (12.2)
Women, N (%)	98 (56)	73 (45)
BMI in kg/m^2, mean (SD)	30.6 (6.5)	27.6 (5.4)
eGFR, median (IQR)	104 (84, 130)	117 (102, 138)
Metformin, N (%)	109 (62)	–
Ethnicity, N (%)		
- Arab	88 (50.3)	100 (61.0)
- Indian	68 (38.9)	37 (22.6)
- Philippines	12 (6.9)	21 (12.8)
- Mixed	7 (4.0)	6 (3.7)

Abbreviations: BMI, body mass index; IQR, interquartile range, N, number of participants, SD, standard deviation; T2D, type 2 diabetes mellitus.

In the sample of 164 controls, we identified 7 protein levels associated with HbA1c (Fig. 2A). Of these, IGFBP-2 was the only protein level that was also observed in the case-control comparison. Notably, a higher Hb1Ac was associated with lower IGFBP-2 levels (−0.50 [SE: 0.10] IGFBP-2 on a Ln2 scale per 1 mmol/L increase in Hb1Ac; p-value = $4.1e^{-6}$). In addition, the lower levels of IGFBP-2 in T2D cases compared to controls were consistently observed across the four different ethnic groups in our study population (Fig. 2B). Additional adjustment of the association between IGFBP-2 and T2D for eGFR did not materially change the study results (results not shown). The partial correlation of IGFBP-2 on T2D was 0.22 (p = $4.1e^{-5}$) in the model adjusted for age, sex, and BMI.

3.3. Metabolomics characterization of IGFBP-2 in controls

Based on the proteomics findings and sensitivity analyses in with Hb1Ac and across the ethnic groups, we took only IGFBP-2 forward to the follow-up metabolomics characterization.

Of the 501 metabolites measured in our study population, 303 were uncorrelated and we therefore used a multiple testing-corrected alpha of $1.7e^{-4}$. In control participants, 20 metabolites were associated with blood IGFBP-2 level, independent of sex, age and body mass index. Of these, 8 were positively associated (e.g., pseudouridine and citrulline) and 12 were negatively associated (e.g., leucine, oxovalerate and hydroxyisobutyrate) with the level of IGFBP-2 in blood (Fig. 3A). The summary statistics of all metabolites in relation to IGFBP-2 in control participants can be found in Supplementary Table 2. Based on the 20 identified metabolites associated with IGFBP-2 and the 20 highest partial correlations (range: 0.20–0.71), all could be mapped in one of the 6 identified subnetworks. (Fig. 3B). However, only 1 metabolite was directly connected to IGFBP-2 in the network analysis, notably citrulline (partial correlation = 0.21). All other networks did not show connections to IGFBP-2; these results remained similar when we allowed more than 20 edges (results not shown).

3.4. Mendelian Randomization analyses on type 2 diabetes mellitus

Of the proteins that were studied, IGFBP-2 was further explored in the Mendelian Randomization analyses on T2D. A total of 16 independent and nonpalindromic SNPs were used as genetic instruments (p-value $< 1e^{-5}$; no SNPs with p-value $< 5e^{-8}$). With IVW, we observed no evidence for a causal association between IGFBP-2 and T2D (odds ratio: 1; 95%CI: 0.96, 1.04; p-value = 0.97). No mutual different results were observed with MR-Egger or MWA, and there were no indications for bias caused by directional pleiotropy.

4. Discussion

We aimed to identify blood protein concentrations in relation to T2D in a Qatari population using a protein-wide approach, and to further characterize our main findings using metabolomics and Mendelian Randomization. We identified 30 proteins that had different mean levels in T2D cases and controls, of which IGFBP-2 level was also associated with Hb1Ac in controls, independent of obesity. In follow-up analyses performed in controls, IGFBP-2 was associated with the levels of 20 metabolites, of which citrulline was directly connected to IGFBP-2 in network analyses. However, in our two-sample Mendelian Randomization analyses, we did not find evidence that IGFBP-2 level was causally associated with the risk of developing T2D. Collectively, our findings suggest IGFBP-2 is a protein associated with the risk of developing T2D and is associated with early glycaemic disturbances,

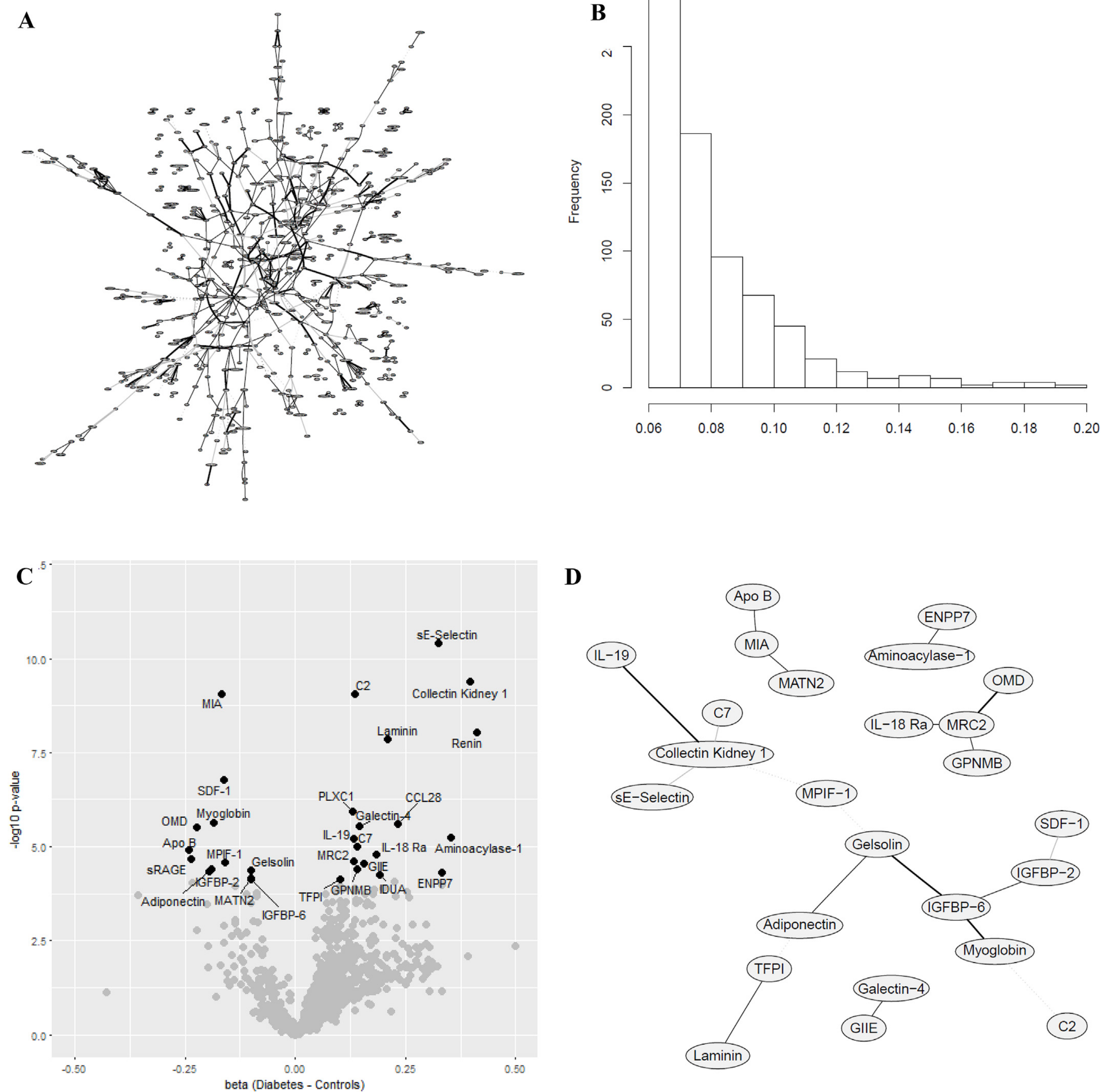


Fig. 1. Differences in protein concentrations between T2D cases and controls.
A) Based on the 1141 proteins measured in the study sample, we selected the 750 strongest partial correlations. These were visualized in the displayed network. B) Distribution of the 750 strongest partial correlations. Partial correlations in the network range between 0.06 and 0.20. C) Differences can be interpreted as the differences between T2D cases and controls on a Ln2-scale. The difference between cases and controls are presented on the x-axis; the -log(p-value) of the statistical comparison in protein concentration between T2D cases and controls is presented on the Y-axis. Black dots are proteins that were statistically significant after correction for multiple testing with Bonferonni based on 584 uncorrelated proteins. Analyses were adjusted for age, sex and body mass index. D) Based on the 30 identified proteins with diabetes mellitus, we selected the 20 strongest partial correlations. This resulted in 5 subnetworks containing at least 2 proteins and covering 22 proteins.

although evidence of causality could not be provided at this point.

The 30 identified proteins that had different mean levels in T2D compared with controls comprised multiple known protein markers. For example, adiponectin has been observed frequently in relation to T2D [25], but causality has not been established [26]. Adiponectin, as well as the closely connected gelsolin, are both secreted by the visceral adipose tissue [27]. Counterintuitively, high concentrations of ApoB have been associated with increased risk of incident T2D, independent of classical lipid risk factors [28], but we found lower concentrations in T2D cases compared to controls. Although we did not have information on treatment adherence, findings could be either influenced by the treatment of the T2D case or because of the comorbidities and complications caused by the T2D. Similarly, the higher concentrations of several of the immune markers (e.g., IL-19), complement factors (e.g.,

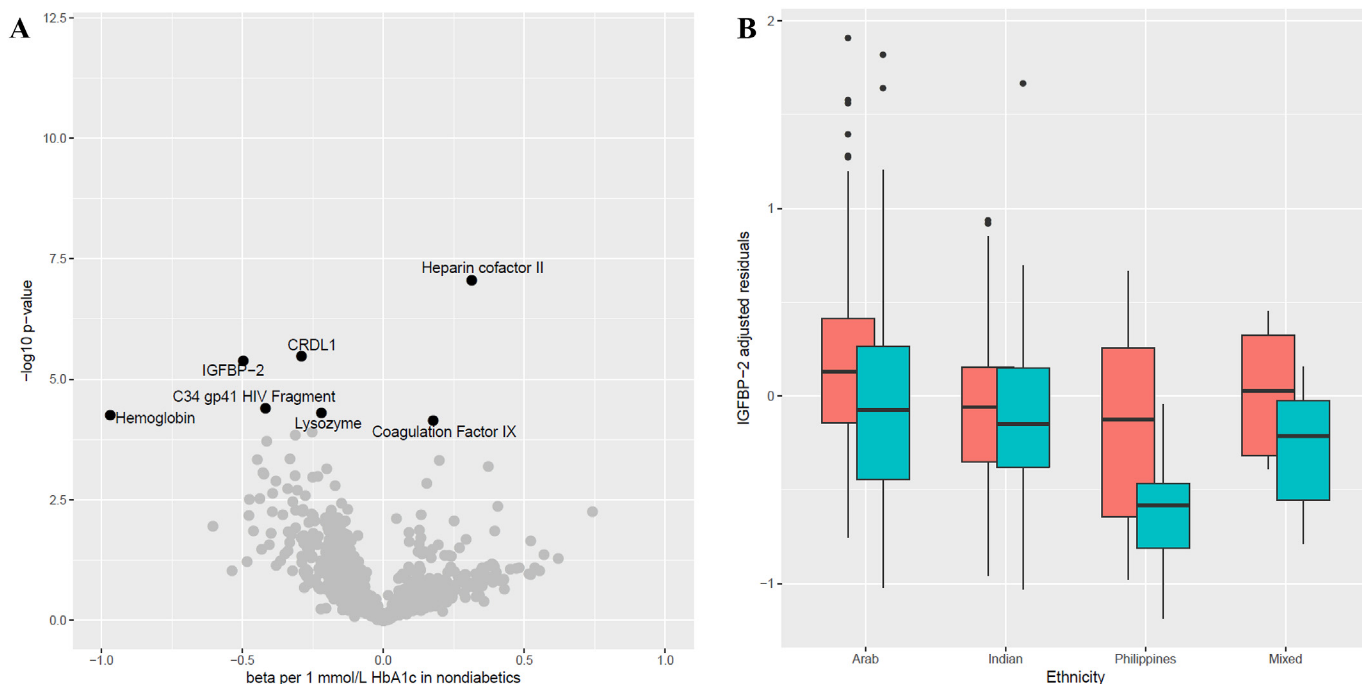


Fig. 2. Follow-up analyses in individuals without diabetes and across ethnic groups.
A) The data presented on the x-axis can be interpreted at the mean difference in protein level on an Ln2 scale per 1 mmol/L increase in HbA1c. the -log(p-value) of the associations is presented on the Y-axis. Black dots are proteins that were statistically significant after correction for multiple testing with Bonferonni based on the number of independent comparisons (notably 584 proteins). Grey dots are proteins that did not reach the level of statistical significance. Analyses were adjusted for age, sex, and body mass index. B) IGFBP-2 levels across the different ethnic groups in the study population are presented as IGFBP-2 residuals adjusted for age, sex and body mass index. Data presented as the median with interquartile range. Controls are presented in red and cases are presented in blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

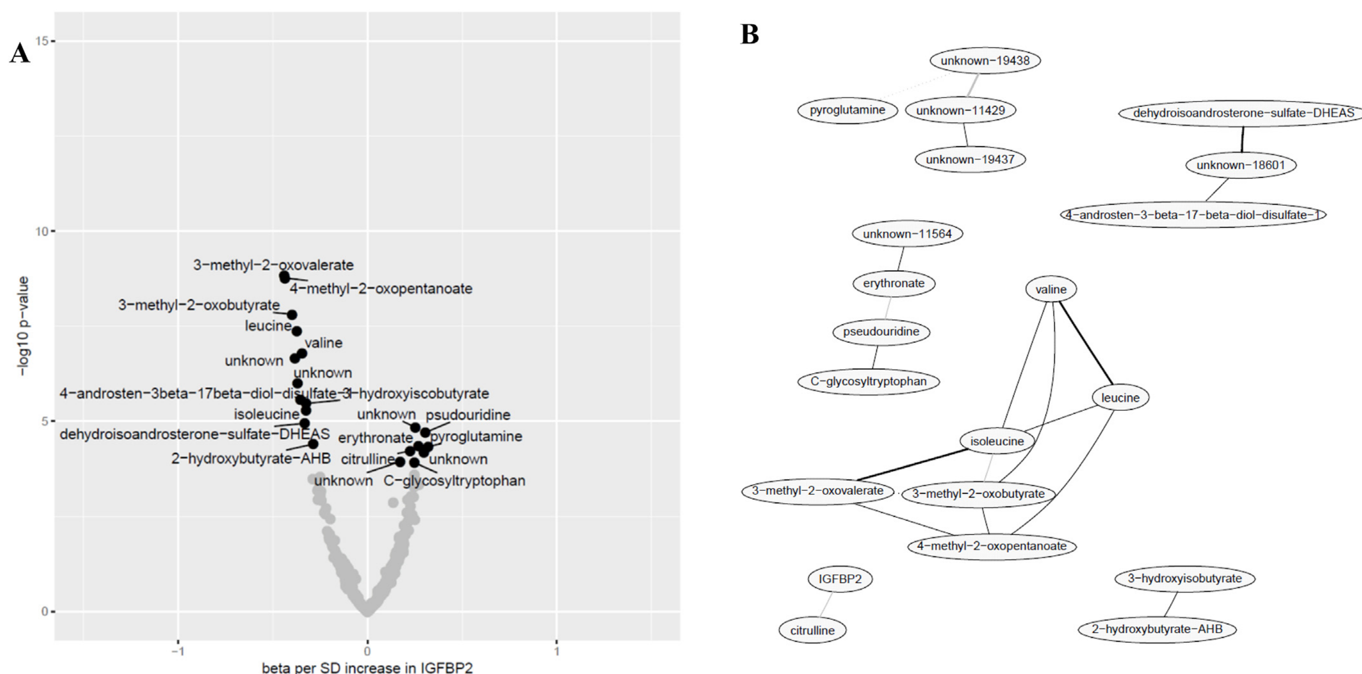


Fig. 3. Association and network analysis between IGFBP-2 level and metabolite levels.
A) The data presented on the x-axis can be interpreted at the mean difference in metabolite level in SD per SD increase in IGFBP-2 level. The -log(p-value) of the associations is presented on the Y-axis. Black dots are proteins that were statistically significant after correction for multiple testing with Bonferonni based on the number of independent comparisons (notably 303 metabolites). Grey dots are proteins that did not reach the level of statistical significance. Analyses were adjusted for age, sex and body mass index. B) Based on the 20 identified metabolites that were associated with IGFBP2 level, we selected the 20 strongest partial correlations. This resulted in 5 subnetworks covering all 20 metabolites including IGFBP2 itself.

C7), and soluble adhesion molecules (e.g., sE-Selectin) could be indicative of a higher cardiovascular risk profile or cardiovascular complications rather than in the causal path in the pathogenesis of T2D. The comparison between cases and controls therefore provided mixed insights of causes and consequences of T2D.

From our protein-wide association analyses, we specifically identified IGFBP-2 as being associated with T2D and early (long term) glycaemic disturbances in individuals without T2D. Recently, IGFBP-2 level has been demonstrated to have an inverse association with T2D, similar as in our study [10,29], as well as with insulin resistance [30]. The biochemical pathways in which IGF binding proteins are involved, notably somatotropic signalling, has been implicated to be involved in insulin signalling and a beneficial ageing process [30,31]. IGFBP-2, as well as the closely connected IGFBP-6, showed specifically high binding affinity to IGF-II [32], although this protein was not present in our used proteomics platform and could not be explored in our study. Previously, IGFBP-2 has been found to be inversely associated with body mass index, and as a potential target for obesity prevention [33–36]. Also, IGFBP-2 has been found to be in part regulated by leptin [35]. We found, independent of body mass index, an association with lower levels of IGFBP-2 in T2D cases. Interestingly, the concentration of IGFBP-2 can be increased by dietary intake as well as by physical exercise [37–39]. Collectively, these results suggest that IGFBP-2 concentrations can be modified using lifestyle interventions, although current evidence is still based on rather extreme conditions. Additional studies are therefore required to increase our understanding of IGFBP-2 in cardiometabolic health and disease in the general population.

In controls, IGFBP-2 was associated with multiple metabolites, among which several have been observed frequently in relation to T2D, such as the branched-chain amino acids (valine and leucine) [40]. However, in our network analyses, we found that IGFBP-2 was specifically connected to the metabolite citrulline. Other metabolites are therefore only associated with IGFBP-2 level, but unlikely to share a similar biological process/pathway. Citrulline is an amino acid that has been previously implicated to enhance hepatic insulin sensitivity [41], and muscle protein synthesis [42]. To the best of our knowledge, there is no literature available further examining the role of IGFBP-2 in citrulline metabolism/signalling. Future research should elucidate the biological meaning of this finding.

Our Mendelian Randomization analyses did not indicate that the association we found between IGFBP-2 and T2D was causal using genetic variants as instruments that have been previously suggestively associated with IGFBP-2 level [5]. It is likely that future GWAS on protein levels will be done with more statistical power, and these are likely to identify more and stronger genetic instruments. Although our Mendelian Randomization analysis was done with suggestive genetic instruments [43], the statistical power is lower and the genetic instruments can contain instruments that are false-positively associated with IGFBP-2. However, it is of interest to note that genetic variation in *IGFBP2* has not been identified in relation with T2D, even not in the most recent European-ancestry genome-wide association study of the DIAGRAM consortium comprising 74,124 T2D cases and 824,006 controls [6]. This results indicates that there is no eQTL for IGFBP-2 present in relation to T2D. Counterintuitively, methylation of the IGFBP-2 gene has been associated with an increased risk of incident T2D [29]. However, it is important to note that DNA methylation is dynamic and similar limitations with respect to causality apply as with the interpretation of conventional epidemiological analyses. Therefore, the exact function of IGFBP-2 in T2D should be further explored in follow-up studies.

The present study had a few limitations to address. The blood samples from the study participants were taken at a random time point during the afternoon. It is currently unknown to what extend proteins measured on our platform exhibit a circadian rhythm or to what extend they are influenced by external factors, such as physical activity and food intake. However, we previously identified robust associations with proteomics and metabolomics within this study sample [4,5,44–46]. Likely, this has resulted in an increased variability in the data and has resulted in reduced statistical power. Proteins with a lower effect size or with strong circadian rhythm have therefore likely been missed in our study. Furthermore, our used case population is heterogenous with respect to treatment (although the majority of our study population took metformin), and disease duration and severity. For example, previous results showed that metformin increases IGFBP-2 mRNA expression levels [35]. Also, to evaluate the prognostic value of IGFBP-2 in clinical practice, prospective studies are warranted [10]. In addition, the Mendelian Randomization analysis that we conducted used genetic instruments that might have been too weak to observe any statistical evidence of a causal association.

In summary, in the present protein-wide association study on T2D, we identified levels of several proteins to be different between T2D cases and controls, independent of obesity. However, only IGFBP-2 was also detectable in relation to early disturbances in glucose levels in controls. Although causality of this finding could not be demonstrated, IGFBP-2 was associated with multiple known T2D-related biochemical pathways. The specific connection with citrulline is of interest in follow-up studies. Therefore, our study provides novel insights in the development of T2D and its progression.

Funding

This work was supported by Biomedical Research Program funds at Weill Cornell Medical College in Qatar, a program funded by the Qatar Foundation. Support for some of the experiments was provided by the Weill Cornell Medical College in Qatar bioinformatics and virtual metabolomics core, which is funded by the Qatar Foundation. DOMK was supported by the Dutch Science Organization (ZonMW-VENI Grant 916.14.023). RN and DvH were supported by the European Commission funded project HUMAN (Health-2013-INNOVATION-1-602757).

Contribution statement

RN, DvH, KS, JK, DOMK: Substantial contributions to conception, design, acquisition of data, analyses and data interpretation; Drafting article and critically comment on the initial versions of the manuscript; Final approval of the manuscript before submission.

Declaration of competing interest

Dennis O Mook-Kanamori works as a part-time clinical research consultant for Metabolon, Inc. All other co-authors declare to have no conflict of interest.

Acknowledgements

The statements made herein are solely the responsibility of the authors. At Weill Cornell Medical College-Qatar, we express our appreciation to Zeinab El-Din and Walaa El Maraghy for their assistance in the data collection, to Abeer Gohar and Sherryl Payra for their support with the administrative questions regarding the project, and to Thomas Doyle for organizing the sample shipment and overseeing laboratory safety. Furthermore, we thank all the support staff at the Department of Dermatology of Hamad Medical Corporation. Most of all, we thank all the study participants for their contribution to this diabetes research study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.abb.2020.108476.

References

[1] Y. Zheng, S.H. Ley, F.B. Hu, Global aetiology and epidemiology of type 2 diabetes mellitus and its complications, Nat. Rev. Endocrinol. 14 (2) (2018) 88–98.

[2] F.B. Hu, J.E. Manson, M.J. Stampfer, G. Colditz, S. Liu, C.G. Solomon, W.C. Willett, Diet, lifestyle, and the risk of type 2 diabetes mellitus in women, N. Engl. J. Med. 345 (11) (2001) 790–797.

[3] J. Tuomilehto, J. Lindstrom, J.G. Eriksson, T.T. Valle, H. Hamalainen, P. Ilanne-Parikka, S. Keinanen-Kiukaanniemi, M. Laakso, A. Louheranta, M. Rastas, V. Salminen, M. Uusitupa, G. Finnish Diabetes Prevention Study, Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance, N. Engl. J. Med. 344 (18) (2001) 1343–1350.

[4] D.O. Mook-Kanamori, M.M. Selim, A.H. Takiddin, H. Al-Homsi, K.A. Al-Mahmoud, A. Al-Obaidli, M.A. Zirie, J. Rowe, N.A. Yousri, E.D. Karoly, T. Kocher, W. Sekkal Gherbi, O.M. Chidiac, M.J. Mook-Kanamori, S. Abdul Kader, W.A. Al Muftah, C. McKeon, K. Suhre, 1,5-Anhydroglucitol in saliva is a noninvasive marker of short-term glycemic control, J. Clin. Endocrinol. Metab. 99 (3) (2014) E479–E483.

[5] K. Suhre, M. Arnold, A.M. Bhagwat, R.J. Cotton, R. Engelke, J. Raffler, H. Sarwath, G. Thareja, A. Wahl, R.K. DeLisle, L. Gold, M. Pezer, G. Lauc, M.A. El-Din Selim, D.O. Mook-Kanamori, E.K. Al-Dous, Y.A. Mohamoud, J. Malek, K. Strauch, H. Grallert, A. Peters, G. Kastenmuller, C. Gieger, J. Graumann, Connecting genetic risk to disease end points through the human blood plasma proteome, Nat. Commun. 8 (2017) 14357.

[6] A. Mahajan, D. Taliun, M. Thurner, N.R. Robertson, J.M. Torres, N.W. Rayner, A.J. Payne, V. Steinthorsdottir, R.A. Scott, N. Grarup, J.P. Cook, E.M. Schmidt, M. Wuttke, C. Sarnowski, R. Magi, J. Nano, C. Gieger, S. Trompet, C. Lecoeur, M.H. Preuss, B.P. Prins, X. Guo, L.F. Bielak, J.E. Below, D.W. Bowden, J.C. Chambers, Y.J. Kim, M.C.Y. Ng, L.E. Petty, X. Sim, W. Zhang, A.J. Bennett, J. Bork-Jensen, C.M. Brummett, M. Canouil, K.U. Ec Kardt, K. Fischer, S.L.R. Kardia, F. Kronenberg, K. Lall, C.T. Liu, A.E. Locke, J. Luan, I. Ntalla, V. Nylander, S. Schonherr, C. Schurmann, L. Yengo, E.P. Bottinger, I. Brandslund, C. Christensen, G. Dedoussis, J.C. Florez, I. Ford, O.H. Franco, T.M. Frayling, V. Giedraitis, S. Hackinger, A.T. Hattersley, C. Herder, M.A. Ikram, M. Ingelsson, M.E. Jorgensen, T. Jorgensen, J. Kriebel, J. Kuusisto, S. Ligthart, C.M. Lindgren, A. Linneberg, V. Lyssenko, V. Mamakou, T. Meitinger, K.L. Mohlke, A.D. Morris, G. Nadkarni, J.S. Pankow, A. Peters, N. Sattar, A. Stancakova, K. Strauch, K.D. Taylor, B. Thorand, G. Thorleifsson, U. Thorsteinsdottir, J. Tuomilehto, D.R. Witte, J. Dupuis, P.A. Peyser, E. Zeggini, R.J.F. Loos, P. Froguel, E. Ingelsson, L. Lind, L. Groop, M. Laakso, F.S. Collins, J.W. Jukema, C.N.A. Palmer, H. Grallert, A. Metspalu, A. Dehghan, A. Kottgen, G.R. Abecasis, J.B. Meigs, J.I. Rotter, J. Marchini, O. Pedersen, T. Hansen, C. Langenberg, N.J. Wareham, K. Stefansson, A.L. Gloyn, A.P. Morris, M. Boehnke, M.I. McCarthy, Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps, Nat. Genet. 50 (11) (2018) 1505–1513.

[7] H. Hwang, B.P. Bowen, N. Lefort, C.R. Flynn, E.A. De Filippis, C. Roberts, C.C. Smoke, C. Meyer, K. Hojlund, Z. Yi, L.J. Mandarino, Proteomics analysis of human skeletal muscle reveals novel abnormalities in obesity and type 2 diabetes, Diabetes 59 (1) (2010) 33–42.

[8] J.C. Rohloff, A.D. Gelinas, T.C. Jarvis, U.A. Ochsner, D.J. Schneider, L. Gold, N. Janjic, Nucleic acid ligands with protein-like side chains: modified aptamers and their use as diagnostic and therapeutic agents, Mol. Ther. Nucleic Acids 3 (2014) e201.

[9] E. Lopez-Villar, G.A. Martos-Moreno, J.A. Chowen, S. Okada, J.J. Kopchick, J. Argente, A proteomic approach to obesity and type 2 diabetes, J. Cell Mol. Med. 19 (7) (2015) 1455–1470.

[10] V. Gudmundsdottir, V. Emilsson, T. Aspelund, M. Ilkov, E.F. Gudmundsson, N.R. Zilhão, J.R. Lamb, L.L. Jennings, V. Gudnason, Deep Serum Proteomics Reveal Biomarkers and Causal Candidates for Type 2 Diabetes, (2019), p. 633297.

[11] D.A. Lawlor, R.M. Harbord, J.A. Sterne, N. Timpson, G. Davey Smith, Mendelian randomization: using genes as instruments for making causal inferences in epidemiology, Stat. Med. 27 (8) (2008) 1133–1163.

[12] B.B. Sun, J.C. Maranville, J.E. Peters, D. Stacey, J.R. Staley, J. Blackshaw, S. Burgess, T. Jiang, E. Paige, P. Surendran, C. Oliver-Williams, M.A. Kamat, B.P. Prins, S.K. Wilcox, E.S. Zimmerman, A. Chi, N. Bansal, S.L. Spain, A.M. Wood, N.W. Morrell, J.R. Bradley, N. Janjic, D.J. Roberts, W.H. Ouwehand, J.A. Todd, N. Soranzo, K. Suhre, D.S. Paul, C.S. Fox, R.M. Plenge, J. Danesh, H. Runz, A.S. Butterworth, Genomic atlas of the human plasma proteome, Nature 558 (7708) (2018) 73–79.

[13] M.A. de Graaf, K.J. Jager, C. Zoccali, F.W. Dekker, Matching, an appealing method to avoid confounding? Nephron Clin. Pract. 118 (4) (2011) c315–c318.

[14] T. Faresjo, A. Faresjo, To match or not to match in epidemiological studies–same outcome but less power, Int. J. Environ. Res. Publ. Health 7 (1) (2010) 325–332.

[15] D.W. Cockcroft, M.H. Gault, Prediction of creatinine clearance from serum creatinine, Nephron 16 (1) (1976) 31–41.

[16] A.M. Evans, C.D. DeHaven, T. Barrett, M. Mitchell, E. Milgram, Integrated, non-targeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems, Anal. Chem. 81 (16) (2009) 6656–6667.

[17] J. Schaefer, R. Opgen-Rhein, K. Strimmer, GeneNet: Modeling and Inferring Gene Networks, (2015) R package version 1.2.13 (available from: https://CRAN.R-project.org/package=GeneNet.

[18] J. Li, L. Ji, Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix, Heredity 95 (3) (2005) 221–227.

[19] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag, New York, 2016.

[20] R. Noordam, R.A. Smit, I. Postmus, S. Trompet, D. van Heemst, Assessment of causality between serum gamma-glutamyltransferase and type 2 diabetes mellitus using publicly available data: a Mendelian randomization study, Int. J. Epidemiol. 45 (6) (2016) 1953–1960.

[21] S. Burgess, R.A. Scott, N.J. Timpson, G. Davey Smith, S.G. Thompson, E.-I. Consortium, Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors, Eur. J. Epidemiol. 30 (7) (2015) 543–552.

[22] J. Bowden, G. Davey Smith, S. Burgess, Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression, Int. J. Epidemiol. 44 (2) (2015) 512–525.

[23] J. Bowden, G. Davey Smith, P.C. Haycock, S. Burgess, Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator, Genet. Epidemiol. 40 (4) (2016) 304–314.

[24] G. Hemani, J. Zheng, B. Elsworth, K.H. Wade, V. Haberland, D. Baird, C. Laurin, S. Burgess, J. Bowden, R. Langdon, V.Y. Tan, J. Yarmolinsky, H.A. Shihab, N.J. Timpson, D.M. Evans, C. Relton, R.M. Martin, G. Davey Smith, T.R. Gaunt, P.C. Haycock, The MR-Base platform supports systematic causal inference across the human phenome, Elife 7 (2018).

[25] S. Li, H.J. Shin, E.L. Ding, R.M. van Dam, Adiponectin levels and risk of type 2 diabetes: a systematic review and meta-analysis, J. Am. Med. Assoc. 302 (2) (2009) 179–188.

[26] H. Yaghootkar, C. Lamina, R.A. Scott, Z. Dastani, M.F. Hivert, L.L. Warren, A. Stancakova, S.G. Buxbaum, L.P. Lyytikainen, P. Henneman, Y. Wu, C.Y. Cheung, J.S. Pankow, A.U. Jackson, S. Gustafsson, J.H. Zhao, C.M. Ballantyne, W. Xie, R.N. Bergman, M. Boehnke, F. el Bouazzaoui, F.S. Collins, S.H. Dunn, J. Dupuis, N.G. Forouhi, C. Gillson, A.T. Hattersley, J. Hong, M. Kahonen, J. Kuusisto, L. Kedenko, F. Kronenberg, A. Doria, T.L. Assimes, E. Ferrannini, T. Hansen, K. Hao, H. Haring, J.W. Knowles, C.M. Lindgren, J.J. Nolan, J. Paananen, O. Pedersen, T. Quertermous, U. Smith, G. Consortium, R. Consortium, T. Lehtimaki, C.T. Liu, R.J. Loos, M.I. McCarthy, A.D. Morris, R.S. Vasan, T.D. Spector, T.M. Teslovich, J. Tuomilehto, K.W. van Dijk, J.S. Viikari, N. Zhu, C. Langenberg, E. Ingelsson, R.K. Semple, A.R. Sinaiko, C.N. Palmer, M. Walker, K.S. Lam, B. Paulweber, K.L. Mohlke, C. van Duijn, O.T. Raitakari, A. Bidulescu, N.J. Wareham, M. Laakso, D.M. Waterworth, D.A. Lawlor, J.B. Meigs, J.B. Richards, T.M. Frayling, Mendelian randomization studies do not support a causal role for reduced circulating adiponectin levels in insulin resistance and type 2 diabetes, Diabetes 62 (10) (2013) 3589–3598.

[27] G. Alvarez-Llamas, E. Szalowska, M.P. de Vries, D. Weening, K. Landman, A. Hoek, B.H. Wolffenbuttel, H. Roelofsen, R.J. Vonk, Characterization of the human visceral adipose tissue secretome, Mol. Cell. Proteomics 6 (4) (2007) 589–600.

[28] S.H. Ley, S.B. Harris, P.W. Connelly, M. Mamakeesick, J. Gittelsohn, T.M. Wolever, R.A. Hegele, B. Zinman, A.J. Hanley, Association of apolipoprotein B with incident type 2 diabetes in an aboriginal Canadian population, Clin. Chem. 56 (4) (2010) 666–670.

[29] C. Wittenbecher, M. Ouni, O. Kuxhaus, M. Jahnert, P. Gottmann, A. Teichmann, K. Meidtner, J. Kriebel, H. Grallert, T. Pischon, H. Boeing, M.B. Schulze, A. Schurmann, Insulin-Like growth factor binding protein 2 (IGFBP-2) and the risk of developing type 2 diabetes, Diabetes 68 (1) (2019) 188–197.

[30] V.C. Russo, W.J. Azar, S.W. Yau, M.A. Sabin, G.A. Werther, IGFBP-2: the dark horse in metabolism and cancer, Cytokine Growth Factor Rev. 26 (3) (2015) 329–346.

[31] A. Bartke, L.Y. Sun, V. Longo, Somatotropic signaling: trade-offs between growth, reproductive development, and longevity, Physiol. Rev. 93 (2) (2013) 571–598.

[32] S.M. Firth, R.C. Baxter, Cellular actions of the insulin-like growth factor binding proteins, Endocr. Rev. 23 (6) (2002) 824–854.

[33] S.B. Wheatcroft, M.T. Kearney, A.M. Shah, V.A. Ezzat, J.R. Miell, M. Modo, S.C. Williams, W.P. Cawthorn, G. Medina-Gomez, A. Vidal-Puig, J.K. Sethi, P.A. Crossey, IGF-binding protein-2 protects against the development of obesity and insulin resistance, Diabetes 56 (2) (2007) 285–294.

[34] A. van den Beld, O. Carlson, M.E. Doyle, D. Rizopoulos, L. Ferrucci, A.J. Van der Lely, J. Egan, IGFBP-2 and aging; A 20 Year longitudinal study on IGFBP-2, IGF-I, BMI, insulin sensitivity and mortality in an aging population, Eur. J. Endocrinol. 180 (2) (2019 Feb 1) 109–116.

[35] K. Hedbacker, K. Birsoy, R.W. Wysocki, E. Asilmaz, R.S. Ahima, I.S. Farooqi, J.M. Friedman, Antidiabetic effects of IGFBP2, a leptin-regulated gene, Cell Metabol. 11 (1) (2010) 11–22.

[36] J. Frystyk, C. Skjaerbaek, E. Vestbo, S. Fisker, H. Orskov, Circulating levels of free insulin-like growth factors in obese subjects: the impact of type 2 diabetes, Diabetes Metab. Res. Rev. 15 (5) (1999) 314–322.

[37] U. Berg, J.K. Enqvist, C.M. Mattsson, C. Carlsson-Skwirut, C.J. Sundberg, B. Ekblom, P. Bang, Lack of sex differences in the IGF-IGFBP response to ultra endurance exercise, Scand. J. Med. Sci. Sports 18 (6) (2008) 706–714.

[38] S.M. Gregory, B.A. Spiering, J.A. Alemany, A.P. Tuckow, K.R. Rarick, J.S. Staab, D.L. Hatfield, W.J. Kraemer, C.M. Maresh, B.C. Nindl, Exercise-induced insulin-like growth factor I system concentrations after training in women, Med. Sci. Sports Exerc. 45 (3) (2013) 420–428.

[39] D.R. Counts, H. Gwirtsman, L.M. Carlsson, M. Lesem, G.B. Cutler Jr., The effect of anorexia nervosa and refeeding on growth hormone-binding protein, the insulin-like growth factors (IGFs), and the IGF-binding proteins, J. Clin. Endocrinol. Metab. 75 (3) (1992) 762–767.

[40] M. Guasch-Ferre, A. Hruby, E. Toledo, C.B. Clish, M.A. Martinez-Gonzalez, J. Salas-Salvado, F.B. Hu, Metabolomics in prediabetes and diabetes: a systematic review and meta-analysis, Diabetes Care 39 (5) (2016) 833–846.

[41] H. Yoshitomi, M. Momoo, X. Ma, Y. Huang, S. Suguro, Y. Yamagishi, M. Gao, L-Citrulline increases hepatic sensitivity to insulin by reducing the phosphorylation of

serine 1101 in insulin receptor substrate-1, BMC Compl. Alternative Med. 15 (2015) 188.
[42] S. Le Plenier, A. Goron, A. Sotiropoulos, E. Archambault, C. Guihenneuc, S. Walrand, J. Salles, M. Jourdan, N. Neveux, L. Cynober, C. Moinard, Citrulline directly modulates muscle protein synthesis via the PI3K/MAPK/4E-BP1 pathway in a malnourished state: evidence from in vivo, ex vivo, and in vitro studies, Am. J. Physiol. Endocrinol. Metab. 312 (1) (2017) E27–E36.
[43] S. Sanna, N.R. van Zuydam, A. Mahajan, A. Kurilshikov, A. Vich Vila, U. Vosa, Z. Mujagic, A.A.M. Masclee, D. Jonkers, M. Oosting, L.A.B. Joosten, M.G. Netea, L. Franke, A. Zhernakova, J. Fu, C. Wijmenga, M.I. McCarthy, Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases, Nat. Genet. 51 (4) (2019) 600–605.
[44] S.B. Zaghlool, D.O. Mook-Kanamori, S. Kader, N. Stephan, A. Halama, R. Engelke, H. Sarwath, E.K. Al-Dous, Y.A. Mohamoud, W. Roemisch-Margl, J. Adamski, G. Kastenmuller, N. Friedrich, A. Visconti, P.C. Tsai, T. Spector, J.T. Bell, M. Falchi, A. Wahl, M. Waldenberger, A. Peters, C. Gieger, M. Pezer, G. Lauc, J. Graumann, J.A. Malek, K. Suhre, Deep molecular phenotypes link complex disorders and physiological insult to CpG methylation, Hum. Mol. Genet. 27 (6) (2018) 1106–1121.
[45] S.B. Zaghlool, B. Kuhnel, M.A. Elhadad, S. Kader, A. Halama, G. Thareja, R. Engelke, H. Sarwath, E.K. Al-Dous, Y.A. Mohamoud, T. Meitinger, R. Wilson, K. Strauch, A. Peters, D.O. Mook-Kanamori, J. Graumann, J.A. Malek, C. Gieger, M. Waldenberger, K. Suhre, Epigenetics meets proteomics in an epigenome-wide association study with circulating blood plasma protein traits, Nat. Commun. 11 (1) (2020) 15.
[46] N.A. Yousri, D.O. Mook-Kanamori, M.M. Selim, A.H. Takiddin, H. Al-Homsi, K.A. Al-Mahmoud, E.D. Karoly, J. Krumsiek, K.T. Do, U. Neumaier, M.J. Mook-Kanamori, J. Rowe, O.M. Chidiac, C. McKeon, W.A. Al Muftah, S.A. Kader, G. Kastenmuller, K. Suhre, A systems view of type 2 diabetes-associated metabolic perturbations in saliva, blood and urine at different timescales of glycaemic control, Diabetologia 58 (8) (2015) 1855–1867.