


# Plasma N-glycome shows continuous deterioration as the diagnosis of insulin resistance approaches

Ana Cvetko <sup>1</sup>, Massimo Mangino,<sup>2,3</sup> Marko Tijardović,<sup>1</sup> Domagoj Kifer,<sup>1</sup> Mario Falchi,<sup>2</sup> Toma Keser,<sup>1</sup> Markus Perola,<sup>4</sup> Tim D Spector,<sup>2</sup> Gordan Lauc,<sup>1,5</sup> Cristina Menni,<sup>2</sup> Olga Gornik<sup>1</sup>

**To cite:** Cvetko A, Mangino M, Tijardović M, *et al*. Plasma N-glycome shows continuous deterioration as the diagnosis of insulin resistance approaches. *BMJ Open Diab Res Care* 2021;**9**:e002263. doi:10.1136/bmjdr-2021-002263

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjdr-2021-002263>).

AC and MM contributed equally. CM and OG contributed equally.

Received 11 March 2021  
Accepted 22 August 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY. Published by BMJ.

<sup>1</sup>University of Zagreb Faculty of Pharmacy and Biochemistry, Zagreb, Croatia

<sup>2</sup>Department of Twin Research and Genetic Epidemiology, King's College London, London, UK

<sup>3</sup>NIHR Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London, UK

<sup>4</sup>National Institute for Health and Welfare, Helsinki, Finland

<sup>5</sup>Genos Glycoscience Research Laboratory, Zagreb, Croatia

**Correspondence to**  
Professor Olga Gornik;  
[ogornik@pharma.unizg.hr](mailto:ogornik@pharma.unizg.hr)

## ABSTRACT

**Introduction** Prediction of type 2 diabetes mellitus (T2DM) and its preceding factors, such as insulin resistance (IR), is of great importance as it may allow delay or prevention of onset of the disease. Plasma protein N-glycome has emerged as a promising predictive biomarker. In a prospective longitudinal study, we included patients with a first diagnosis of impaired glucose metabolism (IR or T2DM) to investigate the N-glycosylation's predictive value years before diabetes development.

**Research design and methods** Plasma protein N-glycome was profiled by hydrophilic interaction ultra-performance liquid chromatography in 534 TwinsUK participants free from disease at baseline. This included 89 participants with incident diagnosis of IR or T2DM during the follow-up period (7.14±3.04 years) whose last sample prior to diagnosis was compared using general linear regression with 445 age-matched unrelated controls. Findings were replicated in an independent cohort. Changes in N-glycome have also been presented in connection with time to diagnosis.

**Results** Eight groups of plasma N-glycans were different between incident IR or T2DM cases and controls ( $p<0.05$ ) after adjusting for multiple testing using Benjamini-Hochberg correction. These differences were noticeable up to 10 years prior to diagnosis and are changing continuously as becoming more expressed toward the diagnosis. The prediction model was built using significant glycan traits, displaying a discriminative performance with an area under the receiver operating characteristic curve of 0.77.

**Conclusions** In addition to previous studies, we showed the diagnostic potential of plasma N-glycome in the prediction of both IR and T2DM development years before the clinical manifestation and indicated the continuous deterioration of N-glycome toward the diagnosis.

## INTRODUCTION

Type 2 diabetes mellitus (T2DM) is one of the most prevalent metabolic diseases in the world, with over 400 million individuals living affected with this chronic disease.<sup>1,2</sup> With multiple cardiovascular, metabolic and even neurological comorbidities developing alongside T2DM,<sup>3</sup> the search for effective medications and therapies would be highly beneficial

## SIGNIFICANCE OF THIS STUDY

### WHAT IS ALREADY KNOWN ABOUT THIS SUBJECT?

- ⇒ Type 2 diabetes mellitus (T2DM) and its risk factors represent a huge public health burden and identifying biomarkers that can predict disease onset would allow prevention or delay of onset of T2DM.
- ⇒ There is a wide array of available T2DM prediction models, many of them are relying on pathophysiological changes of common diabetes biomarkers, which are usually a consequence of an already developed metabolic disorder and are therefore often unable to identify at-risk individuals early enough in the disease development process.

### WHAT ARE THE NEW FINDINGS?

- ⇒ In this study, most of our participants had insulin resistance (IR) as their first confirmed diagnosis of impaired glucose metabolism, and since IR is known as a condition which very commonly precedes the development of T2DM for even a decade, we have now evaluated N-glycans as a predictive biomarker even earlier in life.
- ⇒ Our results also indicated the continuous deterioration of N-glycome toward the diagnosis of impaired glucose metabolism.
- ⇒ Our findings point out the importance of using plasma N-glycans in existing and future T2DM prediction models and risk assessment tools since we have shown that glycans encompass the information body mass index data carry, but also provide additional predictive value.

### HOW MIGHT THESE RESULTS CHANGE THE FOCUS OF RESEARCH OR CLINICAL PRACTICE?

- ⇒ It is still unknown whether these changes are the cause or the consequence of the disease; however, it is now certain that they play an important role in the diabetes development and that the change of glycans within a person could warn about possible disease development and allow both the clinician and the patient to take adequate steps to prevent or delay disease development.

both for the drug-developing companies and the patients.<sup>4</sup> However, T2DM is a chronic, ongoing, incurable disease<sup>5,6</sup> held under

control by medication, but only to prevent further development of comorbidities and to resolve symptoms. Even though some studies presented the possibility of reversing T2DM, such claims are still not clinically proven.<sup>7,8</sup>

Thus, type 2 diabetes and its risk factors represent a huge public health burden and identifying biomarkers that can predict disease onset would allow prevention or delay of onset of T2DM. Early T2DM prediction was mostly based on the recognition of insulin resistance (IR) in the pre-diabetic state.

IR, defined as the inability of insulin to increase glucose uptake and usage in an individual as much as in healthy population,<sup>9</sup> is a long-known major risk factor for the development of T2DM.<sup>10</sup> IR accompanied with abnormally increased body fat is a hallmark of pre-diabetes, which has recently been noted as a highly probable major underlying condition for the development of metabolic syndrome.<sup>11</sup> Moreover, pre-diabetes accompanied with IR and deteriorated  $\beta$  cells accounts for approximately 5%–10% of patients with newly diagnosed T2DM per year.<sup>12</sup>

Many studies investigated both basic IR/T2DM prediction models using standard, accessible variables like blood glucose or insulin levels, as well as improved models upgraded with additional biomarker or body measurement data. A review article from 2012<sup>13</sup> combined multiple diabetes prediction models, basic or improved, in order to assess their predictive and discriminative performance. The authors concluded that the basic models performed well on their own, but additional biomarker data made extended models accomplish better results. The following general worldwide progress of technology, computational models and statistical analyses profusely impacted the field of diabetes prediction by using data mining and machine learning to extensively upgrade and improve prediction tools and software.<sup>14–16</sup> Even though there is a wide array of available T2DM prediction models, many of them are relying on pathophysiological changes of common diabetes biomarkers, which are usually a consequence of an already developed metabolic disorder and are therefore often unable to identify at-risk individuals early enough in the disease development process. Taking this into account, there is still a great need for T2DM predictive biomarkers, such as N-glycans, that could identify individuals at the very start of their metabolic deterioration, while they still appear as healthy.

N-glycans are oligosaccharide structures added to the polypeptide sequence of a protein via the enzymatically mediated and highly regulated process of N-glycosylation.<sup>17</sup> N-glycosylation, a process regulated by a complex network of genes,<sup>18</sup> is different from glycation, a non-enzymatic addition of sugars to proteins.<sup>19</sup> N-glycosylation is among the most prevalent co-translational and post-translational modifications, with the vast majority of eukaryotic proteins being glycosylated.<sup>17</sup> Since it is known that post-translational modifications greatly impact structural and functional features of proteins,<sup>20</sup> it comes as no surprise that N-glycans

were examined in the many physiological and pathophysiological conditions caused by protein diversification, in which they showed to be significant.<sup>21, 22</sup> Recent studies have proposed the idea of glycans as functional effectors in various physiological processes, as well as different disorders and diseases, such as diabetes and obesity-induced IR.<sup>23–26</sup> Importantly, mostly proinflammatory-like changes of both plasma and IgG N-glycome were observed in two separate cross-sectional studies of patients with T2DM.<sup>27, 28</sup> Plasma N-glycome was also identified to have great discriminative power for other types of diabetes, like mature-onset diabetes of the young. Specifically, triantennary sialylated plasma N-glycan with antennary fucose was very successful in extracting individuals with early-onset diabetes with damaging *HNFA1* mutations.<sup>29</sup> Our previous studies on N-glycome in T2DM development showed that individuals with a higher risk of disease due to recorded hyperglycemia during acute illness had increased complexity of plasma N-glycome,<sup>30</sup> possibly reflecting the altered flux of glucose through the hexosamine pathway, which produces uridine diphosphate-N-acetylglucosamine, the substrate for N-linked glycosylation.<sup>24</sup> Other factors such as presence of monosaccharide-nucleotides or presence of enzymes included in the process of glycosylation are important for its heterogeneity,<sup>17</sup> as well as the glycan processing pathway through the Golgi apparatus and structural features of the protein part in glycoprotein, especially near the N-glycosylation site, and many more.<sup>31</sup>

Recently, we have also studied glycans as a tool for cardiometabolic risk assessment and proposed the glycan-based score  $GS_{T2D}$ ,<sup>32</sup> developed on a population with incident T2DM from the European Prospective Investigation into Cancer and Nutrition-Potsdam study. These studies confirmed that plasma N-glycome could be an early predictor of T2DM development. Apart from being associated with various diseases and disorders, plasma N-glycosylation has been thoroughly examined in association with lifestyle factors, such as smoking,<sup>33</sup> exercise,<sup>34</sup> sleep<sup>35</sup> and dietary habits.<sup>36</sup> While smoking influenced increase in complexity of glycans,<sup>33</sup> which is associated with a proinflammatory-like glycosylation profile, moderate to vigorous exercise positively affected the plasma glycosylation profile of older female individuals despite them being diagnosed with metabolic syndrome by decreasing levels of triantennary and tetra-antennary glycans,<sup>34</sup> both hallmarks of complex plasma glycoprofile. Interestingly, higher abundance of biantennary glycan structures with core fucose and one or two sialic acid residues was observed in patients with rapid eye movement sleep behavior disorder when compared with healthy controls.<sup>35</sup> These simpler, low-branched glycan structures are usually decreased in patients with health issues. One study found dietary habits influencing interesting changes in plasma N-glycosylation.<sup>36</sup> Increased sialylation was positively associated with a healthier diet, which is opposite to the commonly observed higher abundance of highly sialylated (trisialylated and tetrasialylated) glycans in individuals with various diseases.

Since IR is a known driving factor for T2DM development and its key feature, in addition to previous findings, this study investigated whether plasma N-glycome is predictive of first clinical diagnosis of impaired glucose metabolism, incident IR or T2DM in 534 individuals (89 incident IR/T2DM cases, 445 controls) from the TwinsUK cohort with independent replication. Furthermore, the relationship between N-glycome and time to diagnosis will be explored.

## RESEARCH DESIGN AND METHODS

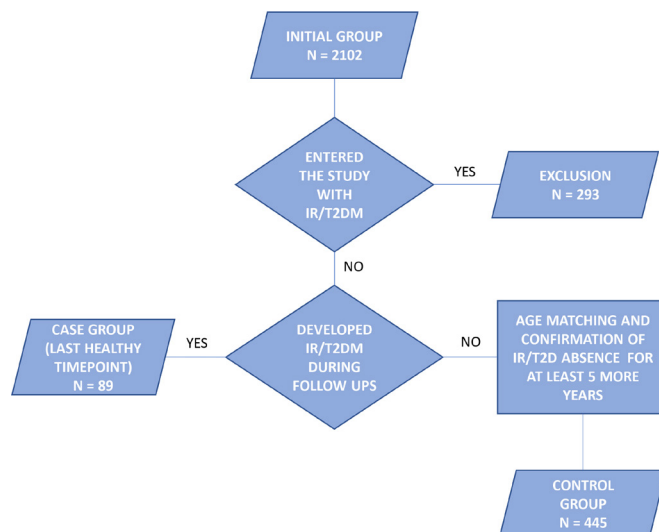
### Study population

#### Discovery cohort

We analyzed 6032 plasma samples from the TwinsUK cohort using the hydrophilic interaction ultra-performance liquid chromatography with fluorescence detection (HILIC-UPLC-FLR). The analyzed samples were collected at multiple timepoints throughout a period of 20 years, from 1996 to 2016. Chromatographic profiling of the total plasma protein N-glycome was followed by the glycan data preprocessing protocol (described in detail in the Statistical analysis section), which reduced the glycan data set to 5889 samples. The following statistical analysis was performed on a subset of 534 women, which included incident IR/T2DM cases and independent controls selected from the original set of samples. Case patients were selected based on their first confirmed occurrence of disruption in glucose metabolism—type 2 diabetes (fasting glucose  $\geq 7$  mmol/L or physician's letter confirming diagnosis) or IR (evaluated using homeostatic model assessment (HOMA2) score)—as well as availability of at least one plasma sample taken before the diagnosis. The control group was selected based on a negative type 2 diabetes and IR status during the total follow-up period and only included samples taken at least 5 years before the last visit, in order to exclude patients who will possibly develop the disease in the next 5 years and to obtain truly negative cases. A flow chart of the inclusion and exclusion protocol is depicted in figure 1. Neither the case nor the control group contained siblings. The control group was age-matched with the case group. Description of the cohort is provided in table 1.

#### Replication cohort

As a replication cohort we used samples collected through the Finland Cardiovascular Risk Study (FINRISK) designed to investigate the risk factors for chronic, non-communicable diseases.<sup>37</sup> We analyzed the samples of 38 participants who developed T2DM (mean age  $\pm$ SD, years=59.42  $\pm$  8.99) during a 10-year follow-up period and 38 participants who remained unaffected and served as controls (mean age  $\pm$ SD, years=59.61  $\pm$  8.93).



**Figure 1** Inclusion flow chart of the TwinsUK cohort. IR, insulin resistance; T2DM, type 2 diabetes mellitus.

#### Consent to participate

All subjects included in this study gave written, informed consent. All twins in the TwinsUK study provided informed written consent.

#### Plasma N-glycome analysis

##### Release of total plasma protein N-glycans

Plasma samples (10  $\mu$ L) were denatured using 20  $\mu$ L of 2% (weight/volume (w/v)) sodium dodecyl sulfate (SDS) (Invitrogen, USA), followed by incubation at 65°C for 10 min. Then, 10  $\mu$ L of 4% (volume/

**Table 1** Descriptive statistics of the analyzed T2DM and IR subset of the discovery TwinsUK cohort

Characteristics	TwinsUK cohort (IR/T2DM subset)	
	Cases (IR/T2DM)	Controls
Samples, n	89 (52/47)	445
Female participants, %	100	100
Age, mean $\pm$ SD, years	55.07 $\pm$ 9.01	55.06 $\pm$ 8.95
BMI, mean $\pm$ SD, kg/m <sup>2</sup>	28.15 $\pm$ 4.78	25.21 $\pm$ 3.99
Follow-up period, mean $\pm$ SD, years	7.14 $\pm$ 3.04	7.14 $\pm$ 3.04
Systolic blood pressure, mean $\pm$ SD, mm Hg	125.73 $\pm$ 16.95	122.81 $\pm$ 16.51
Diastolic blood pressure, mean $\pm$ SD, mm Hg	78.86 $\pm$ 9.91	76.80 $\pm$ 10.12
Active smokers, n (%)	6 (9.4)	15 (5.3)
Non-smokers, n (%)	39 (60.9)	189 (66.3)
Ex-smokers, n (%)	19 (29.7)	81 (28.4)
Data on smoking status were available for 349 participants (65%). BMI, body mass index; IR, insulin resistance; T2DM, type 2 diabetes mellitus.		

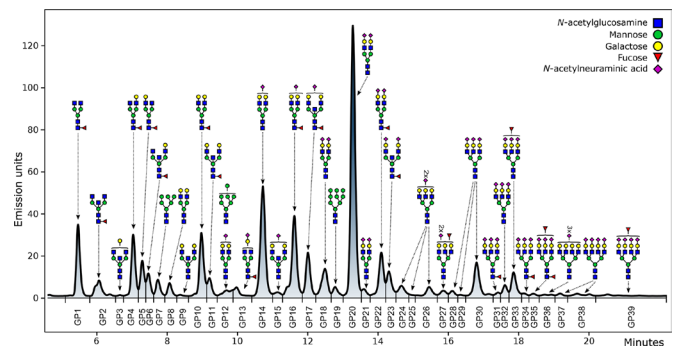
volume (v/v)) Igepal CA-630 (Sigma-Aldrich, St Louis, Missouri, USA) were added to the denatured samples, after which the mixture was shaken for 15 min on a plate shaker (Gujarat Fluorochemicals Limited (GFL), Germany). N-glycans were released by adding 1.2 U of PNGase F (Promega, USA) to the mixture, followed by overnight incubation at 37°C.

### Labeling and HILIC solid-phase extraction clean-up of released N-glycans

N-glycans released from total plasma proteins were labeled with 2-aminobenzamide (2-AB), a fluorescent dye. The fluorescent labeling mixture was made inhouse from both 2-AB (19.2 mg/mL; Sigma-Aldrich) and 2-picoline borane (44.8 mg/mL; Sigma-Aldrich) dissolved in dimethyl sulfoxide (Sigma-Aldrich) and glacial acetic acid (Merck, Germany) mixture (70:30 v/v). Each sample was labeled by addition of 25 µL of the labeling mixture, followed by 2-hour incubation at 65°C. Following the 2-hour incubation, samples were brought to a total concentration of 96% acetonitrile by addition of 700 µL of 100% cold acetonitrile to each sample. The samples were then transferred to a 0.2 µm hydrophilic polypropylene membrane (GHP) filter plate (Pall Corporation, USA). All wells of the filter plate were prewashed with 70% ethanol (Sigma-Aldrich) and water, followed by equilibration using 96% acetonitrile. Both 70% ethanol and 96% acetonitrile were freshly prepared on the day of the experiment execution. Removal of solvent and impurities from samples, such as excess free labeling dye and reducing agent, was performed by HILIC-solid-phase extraction on the mentioned filter plate with the application of vacuum using a vacuum manifold (Millipore Corporation, USA). Loaded samples were subsequently washed 5× with 96% acetonitrile (ACN). Plasma protein released and labeled N-glycans were finally eluted 2× with 90 µL of ultra-pure water and stored at -20°C until usage.

### HILIC-UPLC-FLR profiling of total plasma protein released N-glycans

Fluorescently labeled N-glycans were separated by HILIC on Acquity UPLC H-Class instrument (Waters, Milford, USA) consisting of a quaternary solvent manager, sample manager and a fluorescence detector, set with excitation and emission wavelengths of 250 nm and 428 nm, respectively. The instrument was under the control of Empower V.3 software, build 3471 (Waters). Waters BEH Glycan chromatography column was used to separate labeled N-glycans, with 100 mM ammonium formate, pH 4.4, as solvent A and liquid chromatography-mass spectrometry grade acetonitrile as solvent B. Each 96-well plate contained five standard samples and one blank sample for the purpose of maintaining quality control and performing batch correction. The plasma protein N-glycans separation method used linear gradient of 70%–53% acetonitrile



**Figure 2** Chromatogram of HILIC-UPLC-FLR-analyzed plasma N-glycome. The most abundant glycan structure is graphically presented for each glycan peak (GP). The y axis represents the intensity of the signal measured in emission units, while retention time in minutes is presented on the x axis. HILIC-UPLC-FLR, hydrophilic interaction ultra-performance liquid chromatography with fluorescence detection.

at a flow rate of 0.561 mL/min in a 25 min analytical run. The calibration of the system was done by using an external standard of hydrolyzed and 2-AB labeled glucose oligomers from which the retention times for the individual glycans were converted to glucose units. Automated integration method was used to perform data processing.<sup>38</sup> The plasma proteins N-glycans chromatograms were all separated equivalently into 39 peaks (GP1–GP39) and are presented in figure 2.

Detailed description of glycan structures corresponding to each glycan peak is presented in online supplemental table 1. Glycan peaks were analyzed based on their elution positions and measured in glucose units. The results were then compared with the reference values from the ‘GlycoStore’ database (available at <https://glycostore.org/>)<sup>39</sup> for structure assignment. In addition to 39 directly measured plasma glycan traits, 16 derived traits were calculated. These derived traits represent an average of the particular glycosylation structural feature, like branching, sialylation, galactosylation, fucosylation and incidence of bisecting N-acetylglucosamine (GlcNAc). The formulas used for calculation of plasma protein N-glycan derived traits are presented in online supplemental table 2.

### Statistical analysis

#### Data normalization

The raw data represented as areas under each glycan peak had to be normalized in order to remove the experimental noise. This area was divided by the total integrated area and multiplied by 100, expressing each peak as a percentage of the total integrated area. This allows different patients’ glycan peaks to be compared, regardless of absolute intensities obtained by analysis of their samples.

#### Batch correction

To remove the effects of experimental variation during sample preparation and analysis, batch correction was

performed on normalized glycan data. The data were first log-transformed due to the right-skewness of its distribution. The experimental noise was then reduced by applying the ComBat method from the R package 'sva'. Within the model, the corresponding order number of sample plates, representing the source of variation, was set as a batch covariate. This procedure was performed for every glycan peak.

### Analysis of last prediagnostic timepoint

In order to identify glycan structures which are significantly different in patients prior to diagnosis of T2DM or IR compared with controls, general linear models were used. For each glycan, log-transformed relative area under the glycan peak was modeled as a dependent variable, while disease status was used as an independent variable. Since the control group was age-matched and there were no siblings in either of the groups, no additional covariates were included in the model. This was repeated for each of 39 glycan peaks and 16 derived glycan traits. We adjusted for multiple testing using Benjamini-Hochberg false discovery rate, with adjusted p value of <0.05 considered as significant. Obtained effect sizes represent the natural logarithm of the difference in relative area of the corresponding glycan peak in prediagnostic patients compared with controls.

### Prediction of disease development

Discriminatory potential of individual glycan peaks was further evaluated using area under the receiver operating characteristic curve (AUC). From the set of eight glycan peaks significantly associated with disease status in the previous step, the best predicting ones were identified using stepwise regression variable selection method. It was carried out in bidirectional approach based on Akaike information criteria to find the best performing model and reduce overfitting. Six glycan peaks selected by stepwise regression along with the body mass index (BMI) data were included in the logistic regression model. Using receiving operator characteristic (ROC) curves, it was compared with BMI data alone, which represented the null model, as well as the model containing only six selected glycan peaks. The 'pROC' package was used to construct the ROC curves and calculate the corresponding AUC.

### Temporal progression of glycan abundances

The dynamics of glycan abundances in the disease development period were also represented. All available case group samples, ranging from 10 years before the diagnosis of IR or T2DM up to 2 years after the diagnosis, were split into four groups (10–8 years, 7–5 years, 4–2 years before and 0–2 years after the diagnosis) and included in the linear mixed effects model. As fixed effects, log-transformed relative area under the glycan peak was modeled as the dependent variable, while the corresponding temporal groups along with age at time of sampling were included as independent variables.

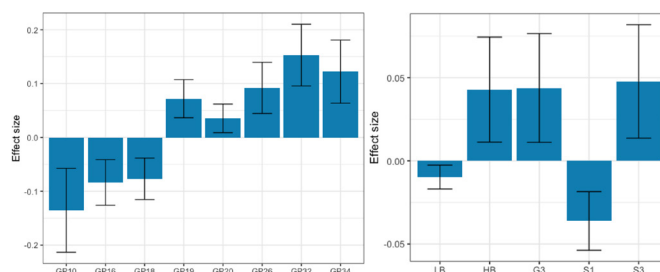
Individual patient identification nested within the family identification was modeled as random effect. From the obtained model using the 'emmeans' package, the mean of glycan peak relative area was estimated for each temporal group and back-transformed from the log scale. This provided age-corrected data on average abundance of individual glycan peak, depending on the time to or from diagnosis. The procedure was repeated for each glycan peak previously identified as significantly changed in the case group compared with controls. Estimated means were then graphically represented along with 95% CI.

## RESULTS

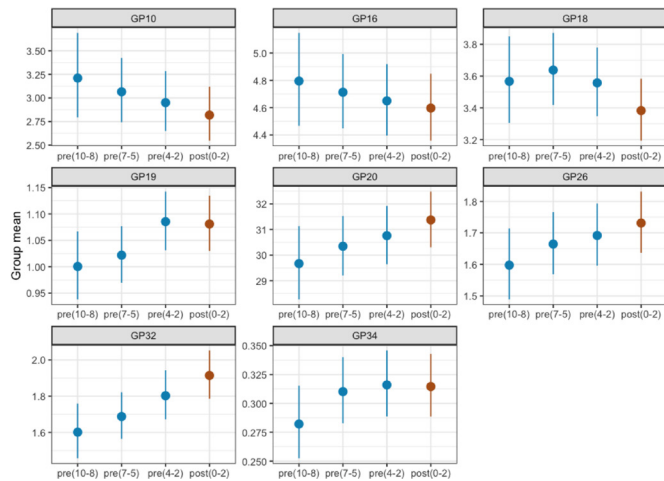
### Plasma N-glycome is extensively different in individuals prior to their IR/T2DM diagnosis when compared with unaffected controls

Using HILIC-UPLC method, we have analyzed and profiled total plasma protein N-glycome in 534 patients from the TwinsUK cohort. Of these, 445 were unaffected individuals who have not developed IR nor T2DM during the study period (controls) and 89 were individuals who were diagnosed with IR/T2DM in one of the timepoints (the IR/T2DM cases). General linear model was used to examine differences in initial and derived plasma glycan traits between the IR/T2DM cases (last sample prior to the disease diagnosis; mean time before diagnosis  $4.65 \pm 2.36$ ) and the controls. Out of 39 directly measured initial glycan traits, 8 showed statistically significant difference. GP10, GP16 and GP18 were decreased, while GP19, GP20, GP26, GP32 and GP34 were increased, in IR/T2DM cases, as shown in **figure 3** and online supplemental table 3.

Besides investigating differences in directly measured initial glycans, we have also searched for significant differences in derived traits, which represent an average of the same glycosylation feature, shared among different glycan structures and are calculated from



**Figure 3** Effects of insulin resistance and type 2 diabetes development on the identified significantly differentiated initial (left) and derived plasma N-glycan traits (right) from the TwinsUK cohort. Calculated effect size (natural logarithm of difference in relative area) of each presented initial or derived trait is shown on the y axis with error bar representing 95% CI, while glycan trait name, initial or derived, is displayed on the x axis. G3, trigalactosylated glycans; GP, glycan peak; HB, high-branched glycans; LB, low-branched glycans; S1, monosialylated glycans; S3, trisialylated glycans.



**Figure 4** Time-to-diagnosis behavior of the most significantly altered initial glycan traits in individuals who developed insulin resistance/type 2 diabetes from the TwinsUK cohort. Glycan abundance is presented on the y axis with dots representing calculated means and lines representing 95% CI, while different temporal groups are presented on the x axis. The timepoint period in which the disease diagnosis occurred is labeled brown, while all prediagnosis timepoint periods are labeled blue. Numbers listed inside the brackets represent years of distance from diagnosis. GP, glycan peak.

the initial glycan traits. After adjusting for multiple testing, out of 16 derived plasma N-glycome traits, 5 were significantly altered in affected individuals when compared with the controls: monosialylated (S1) and low-branched N-glycans showed a significant decrease, while high-branched, trisialosylated (G3) and trisialylated (S3) N-glycans were significantly increased, in IR/T2DM cases (figure 3, online supplemental table 3).

The effects of T2DM and IR development on the identified initial and derived plasma N-glycan traits are depicted separately in online supplemental figure 1.

### Plasma N-glycome continuously changes toward the diagnosis of IR or T2DM

Next, we examined the time-to-diagnosis behavior of eight initial glycan groups that showed predictive potential for IR/T2DM development. Their abundances were plotted through different timepoints from the 10 years prior to disease development to the disease onset (figure 4). Each temporal group consisted of all case samples available for chosen timepoint prior to disease diagnosis; for example, temporal group 8–10 years prior to diagnosis consisted of glycan data analyzed from all available samples taken 8–10 years prior to diagnosis. The groups did not differ in age (the comparison of N-glycan levels with the control group of the corresponding age for each timepoint is shown in online supplemental figure 2, with average age of both cases and controls shown in online supplemental table 4). The differences between temporal groups for GP10, GP16, GP20, GP26 and GP32 showed a continuous trend toward disease development.

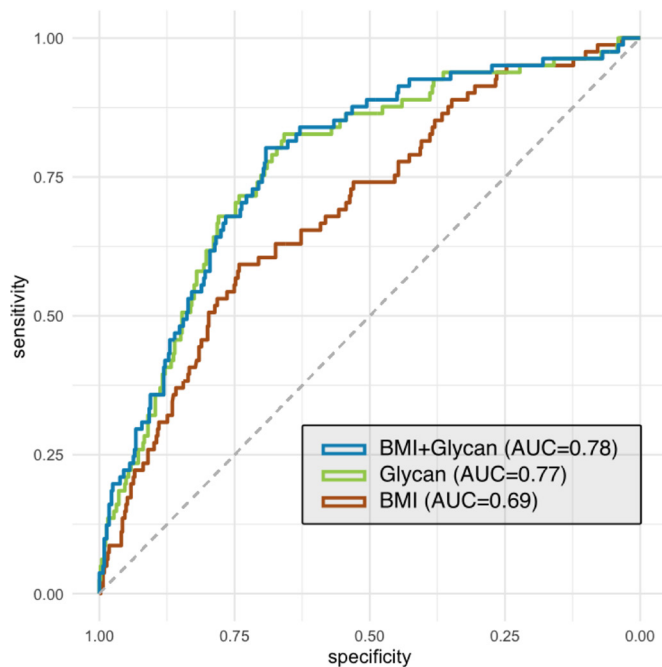
GP19 even shows an increase in this trend 2–4 years prior to diagnosis, while GP18 displays a change in direction 5–7 years before diagnosis establishment. GP34 shows the largest glycan abundance increase from 8–10 to 5–7 years prior to diagnosis, after which their levels start to stagnate. These results demonstrate that the human plasma N-glycome develops more visible and notable alterations as the diagnosis of either IR or T2DM approaches.

### Stratification of IR/T2DM-prone individuals using N-glycans

T2DM prediction and progression models are key factors for diabetes management and could be the source of significant improvements in health status and overall quality of life for the individuals with increased risk for this serious health burden. Therefore, in order to improve the IR/T2DM prediction, we have built a ‘glyco diabetes prediction model’ using the most informative glycan variables (GP10, GP16, GP18, GP19, GP20 and GP34). The model was built from the last available data before disease diagnosis. We further added BMI to the model as it is a known risk factor for IR/T2DM,<sup>40</sup> thus creating a model combining glycan and BMI information (‘glyco-BMI diabetes prediction model’). The performance of each model was examined using ROC curve analysis. The result of ROC analysis is the graphic representation of the stratification ability based on the comparison of glycan abundances in 534 individuals. The AUC value of ‘glyco diabetes prediction model’ is 0.77, while the AUC of the ‘glyco-BMI diabetes prediction model’ is 0.78, and for the model consisting only of BMI data (‘BMI diabetes prediction model’) it is 0.69. These results show that glycans are valuable predictors of IR/T2DM. Furthermore, the fact that BMI has negligible contribution to glycan prediction model indicates that glycans already encompass the information the BMI data can provide (figure 5). Additional data for each point of the ROC curves are provided in online supplemental table 5. We have also compared the predictiveness of glycans with BMI and other risk factor data (smoking status and blood pressure) available only for the portion of TwinsUK cohort participants (online supplemental figure 3).

### External validation of significantly altered initial glycan traits on an independent T2DM follow-up cohort (FINRISK)

We replicated our findings in FINRISK, an independent T2DM cohort. We compared baseline plasma protein N-glycan abundances of both 38 incident cases (the ones who developed T2DM during a 10-year-follow-up) and 38 controls. We were able to show the same direction of changes for all previously identified significantly altered initial N-glycans (GP10, GP18, GP19, GP20, GP26, GP32, GP34), except for GP16, with GP32 displaying statistically significant difference between the groups (adjusted  $p=6.62 \times 10^{-03}$ ; online supplemental table 6). Mentioned glycan alterations in FINRISK cohort are depicted in online supplemental figure 4.



**Figure 5** Stratification performance of insulin resistance/type 2 diabetes prediction model created from the TwinsUK cohort data. Comparison of prediction model based on selected glycan peaks with the model including the BMI data alone and the model combining both BMI and glycan data. Dashed grey diagonal line represents no-discrimination reference. AUC, area under the receiver operating characteristic curve; BMI, body mass index.

## CONCLUSIONS

Here we show that plasma N-glycome can identify individuals who will develop either IR or T2DM years before the clinical diagnosis. Individuals healthy at baseline and developing IR/T2DM during the follow-up time have a higher presence of complex, highly branched glycans within their plasma N-glycome, accompanied by a decrease in low-branched structures. Increase in the complexity and branching of the plasma N-glycome is a hallmark of many pathological conditions with inflammatory component,<sup>41</sup> as well as aging,<sup>21</sup> and was also observed in previous studies investigating the changes of plasma N-glycome connected to T2DM.<sup>27–30</sup> It has been suggested that the increased branching of plasma N-glycome actually reflects the disruption in the glucose metabolism through the altered flux of glucose through the hexosamine pathway.<sup>30</sup> We have previously reported that N-glycans (GS<sub>T2D</sub> score) could predict T2DM development 6–8 years before the disease onset. Herein we strengthen our result by including not only the individuals developing T2DM, but also individuals developing IR as their first diagnosis of impaired glucose metabolism. Since IR is known as a condition which very commonly precedes the development of T2DM for even a decade,<sup>42–43</sup> we have now evaluated the predictive potential of N-glycans earlier in life.

The two glycan peaks which showed the most significant alteration in incident IR/T2DM cases in this study were

GP32 ( $p=9.52\times 10^{-6}$ ) and GP34 ( $p=9.38\times 10^{-4}$ ). GP32 is a structure consisting of a trigalactosylated trisialylated triantennary glycan structure (A3G3S3), while GP34 is the mixture of mostly core fucosylated trigalactosylated trisialylated triantennary glycan structure (FA3G3S3) and less tetragalactosylated trisialylated tetra-antennary glycan structure (A4G4S3). A3G3A3 structure, in plasma, predominantly originates from  $\alpha$ 1-acid glycoprotein (AGP), while A4G4S3 completely derives from this positive acute phase protein,<sup>44</sup> whose increased levels were found to be significantly associated with diabetes in many studies.<sup>45–47</sup> Even in our recent study, which focused on a cardiometabolic risk assessment, GP32 (A3G3S3) was recognized as one of the six structures strongly associated with T2DM development.<sup>32</sup> Considering these findings, future studies should definitely focus on investigating the N-glycosylation of AGP in diabetes, such as a recently published paper which developed high-throughput method for AGP N-glycosylation analysis and tested it on a population of individuals with hyperglycemia at high risk of T2DM development.<sup>48</sup>

Our examination of time-to-diagnosis behavior of significantly altered glycan structures showed that modification of plasma N-glycome started even 10 years before the clinical presentation and diagnosis of glucose metabolism impairment, either IR or T2DM, and it continuously proceeds throughout the disease onset.

To determine the discriminative power of plasma N-glycans in identifying individuals at risk of IR/T2DM, we have built a ‘glyco diabetes prediction model’ using the most significantly different and informative N-glycan structures. This glyco model had the discriminative power greater than the model with BMI data alone and almost identical as the BMI–glycan combined model. Our ‘glyco diabetes prediction model’ had similar performance as other, non-glycan-based, suggested T2DM prediction models.<sup>49</sup> Its clinical value could be possibly further improved when combined with other known risk factors, but unfortunately we did not have sufficient data available for our patients to evaluate this potential. Our findings on diabetes prediction using glycans, as well as other similar studies on this subject,<sup>30–31</sup> definitely point out the importance of using plasma N-glycans in existing and future T2DM prediction models and risk assessment tools since we have shown that glycans encompass the information BMI data carry, but also provide additional predictive value.

Comparison of baseline plasma N-glycans in FINRISK incident diabetes cases with that of healthy controls confirmed our findings on the increase of more complex glycan structures present in plasma N-glycome prior to disease diagnosis. It is important to highlight that in the FINRISK cohort we have compared glycans from individuals’ samples taken 10 years before the T2DM diagnosis occurred, which means that probably some participants in the mean time (baseline timepoint to follow-up timepoint) developed/were diagnosed with IR or pre-diabetes. Therefore, these baseline samples of future

T2DM cases could display the earliest noticeable changes in glycans resulting from underlying development of glucose metabolism impairment.

We note some study limitations. First, the results were discovered primarily in women. Second, our replication cohort is quite small-numbered, which may pose as a risk of bias. Third, due to unavailability of various patient clinical data, we were unable to further test glycans' clinical value by evaluating the discriminative potential of a model composed of glycans and other T2DM risk factors (apart from BMI). Also, it is important to mention that BMI, while widely and most commonly used, is not the best clinical marker for diabetes prediction, especially in older population,<sup>50</sup> where it has been shown that the association of BMI and diabetes decreases with age. On the other hand, BMI is calculated solely from body weight and height and offers no information regarding body fat percentage, which showed to bring added informative value in identifying individuals at high risk of abnormal glucose levels.<sup>51</sup> Unfortunately, body fat percentage data were not available for this study. Fourth, in our study we have investigated only the glycan portion of the plasma glycoproteins, and not the protein part. Changes in protein levels,<sup>52</sup> their clearance rate, structure<sup>53</sup> and other parameters are known to be influenced by a myriad of various processes, such as different diseases,<sup>54</sup> lifestyle habits, oxidative status and others,<sup>55</sup> all of which also heavily influence glycans.<sup>17 21 56</sup> Therefore, it is important to mention that some of the observed glycan alterations might also be a result of changes in the protein portion of plasma glycoproteins.

In conclusion, our study further highlights the role of T2DM prediction using plasma N-glycome. This opens the possibility of early assessment of individuals at high risk of disrupted glucose metabolism which can lead to prevention or delay of T2DM development, increasing the quality of life and at the same time decreasing treatment costs. We have shown that plasma N-glycome is not only altered years before T2DM development, but also precedes IR diagnosis, showing that the impairment of glycome starts to manifest even earlier than it was previously shown. Taking into consideration this early manifestation, glycome changes are certainly associated with underlying progression of T2DM, which is known to take years to develop and present symptoms.<sup>57 58</sup> It is still unknown whether these changes are the cause or the consequence of the disease; however, it is now certain that they play an important role in diabetes development and that the change of glycans within a person could warn about possible disease development and allow both the clinician and the patient to take adequate steps to prevent or delay disease development.

**Acknowledgements** We thank all the participants for contributing and supporting our research.

**Contributors** AC, MT, TK, CM and OG conceived and designed the study. CM and OG supervised the study. AC, MM, MT, DK, MF, TK, MP, TDS, GL, CM and OG participated in data acquisition, collection, analysis or interpretation. AC, MT and

OG drafted the manuscript. AC, MM, MT, DK, MF, TK, MP, TDS, GL, CM and OG critically revised the manuscript for intellectual content. All authors approved the final version of the manuscript. OG is the guarantor of the study and is responsible for the integrity of the work as a whole.

**Funding** This work was supported by the Medical Research Council AimHy (MR/M016560/1) project grant and by the Chronic Disease Research Foundation (grant numbers CDRF-18/2019 and CDRF-17/2019). TwinsUK receives funding from the Wellcome Trust (grant number 212904/Z/18/Z) and from European Commission H2020 grants SYSCID (contract #733100); and the National Institute for Health Research (NIHR) Clinical Research Facility and the Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust, in partnership with King's College London. CM is funded by the Chronic Disease Research Foundation and by the Medical Research Council. MM is funded by the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust, in partnership with King's College London. This work was supported by the European Structural and Investment funding for the 'Croatian National Centre of Research Excellence in Personalized Healthcare' (contract #KK.01.1.1.01.0010), 'Centre of Competences in Molecular Diagnostics' (contract #KK.01.2.2.03.0006), and the European Regional Development Fund grant 'CardioMetabolic' agreement (#KK.01.2.1.02.0321).

**Competing interests** GL declares that he is a founder and owner of Genos Glycoscience, which offers commercial service of glycomic analysis and has several patents in this field.

**Patient consent for publication** Not required.

**Ethics approval** The TwinsUK study was approved by NRES Committee London - Westminster, under referent number EC04/015. The FINRISK study was approved by the Ethics Committee of Helsinki and Uusimaa Hospital District, under referent numbers 558/E3/2001 and 229/E0/06. Both studies were conducted according to the Declaration of Helsinki.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request. The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

#### ORCID iD

Ana Cvetko <http://orcid.org/0000-0002-7134-2484>

#### REFERENCES

- 1 IDF atlas 9th edition and other resources. Available: <https://www.diabetesatlas.org/en/resources/> [Accessed 30 Apr 2020].
- 2 Saeedi P, Petersohn I, Salpea P, *et al.* Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9<sup>th</sup> edition. *Diabetes Res Clin Pract* 2019;157:107843.
- 3 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.
- 4 Shomali M. Diabetes treatment in 2025: can scientific advances keep pace with prevalence? *Ther Adv Endocrinol Metab* 2012;3:163–73.
- 5 Olokoba AB, Obateru OA, Olokoba LB. Type 2 diabetes mellitus: a review of current trends. *Oman Med J* 2012;27:269–73.

- 6 Diabetes. Available: <https://www.who.int/westernpacific/health-topics/diabetes> [Accessed 30 Apr 2020].
- 7 Hallberg SJ, Gershuni VM, Hazbun TL, et al. Reversing type 2 diabetes: a narrative review of the evidence. *Nutrients* 2019;11:766.
- 8 Amer OE, Sabico S, Alfawaz HA, et al. Reversal of prediabetes in Saudi adults: results from an 18 month lifestyle intervention. *Nutrients* 2020;12:804.
- 9 Lebovitz HE. Insulin resistance: definition and consequences. *Exp Clin Endocrinol Diabetes* 2001;109 Suppl 2:S135–48.
- 10 Lillioja S, Mott DM, Spraul M, et al. Insulin resistance and insulin secretory dysfunction as precursors of non-insulin-dependent diabetes mellitus. prospective studies of Pima Indians. *N Engl J Med* 1993;329:1988–92.
- 11 Mayans L. Metabolic syndrome: insulin resistance and prediabetes. *FP Essent* 2015;435:11–16.
- 12 Tabák AG, Herder C, Rathmann W, et al. Prediabetes: a high-risk state for diabetes development. *Lancet* 2012;379:2279–90.
- 13 Abbasi A, Peelen LM, Corpeleijn E, et al. Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study. *BMJ* 2012;345:e5900.
- 14 Razavian N, Blecker S, Schmidt AM, et al. Population-Level prediction of type 2 diabetes from claims data and analysis of risk factors. *Big Data* 2015;3:277–87.
- 15 Wu H, Yang S, Huang Z, et al. Type 2 diabetes mellitus prediction model based on data mining. *Inform Med Unlocked* 2018;10:100–7.
- 16 Zou Q, Qu K, Luo Y, et al. Predicting diabetes mellitus with machine learning techniques. *Front Genet* 2018;9:515.
- 17 Varki A, Cummings RD, Esko JD, eds. *Essentials of Glycobiology*. 3rd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2015.
- 18 Lauc G, Rudan I, Campbell H, et al. Complex genetic regulation of protein glycosylation. *Mol Biosyst* 2010;6:329–35.
- 19 Yan SF, Ramasamy R, Naka Y, et al. Glycation, inflammation, and RAGE: a scaffold for the macrovascular complications of diabetes and beyond. *Circ Res* 2003;93:1159–69.
- 20 Mann M, Jensen ON. Proteomic analysis of post-translational modifications. *Nat Biotechnol* 2003;21:255–61.
- 21 Dotz V, Wuhler M. N-Glycome signatures in human plasma: associations with physiology and major diseases. *FEBS Lett* 2019;593:2966–76.
- 22 Gudelj I, Lauc G, Pezer M. Immunoglobulin G glycosylation in aging and diseases. *Cell Immunol* 2018;333:65–79.
- 23 Abdel Rahman AM, Ryczko M, Nakano M, et al. Golgi N-glycan branching N-acetylglucosaminyltransferases I, V and VI promote nutrient uptake and metabolism. *Glycobiology* 2015;25:225–40.
- 24 Ohtsubo K, Chen MZ, Olefsky JM, et al. Pathway to diabetes through attenuation of pancreatic beta cell glycosylation and glucose transport. *Nat Med* 2011;17:1067–75.
- 25 Tanigaki K, Sacharidou A, Peng J, et al. Hyposialylated IgG activates endothelial IgG receptor FcγRIIB to promote obesity-induced insulin resistance. *J Clin Invest* 2018;128:309–22.
- 26 Ohtsubo K, Takamatsu S, Gao C, et al. N-Glycosylation modulates the membrane sub-domain distribution and activity of glucose transporter 2 in pancreatic beta cells. *Biochem Biophys Res Commun* 2013;434:346–51.
- 27 Dotz V, Lemmers RFH, Reiding KR, et al. Plasma protein N-glycan signatures of type 2 diabetes. *Biochim Biophys Acta Gen Subj* 2018;1862:2613–22.
- 28 Li X, Wang H, Russell A, et al. Type 2 diabetes mellitus is associated with the immunoglobulin G N-glycome through putative proinflammatory mechanisms in an Australian population. *OMICS* 2019;23:631–9.
- 29 Juszczak A, Pavić T, Vučković F, et al. Plasma fucosylated glycans and C-reactive protein as biomarkers of HNF1A-MODY in young adult-onset nonautoimmune diabetes. *Diabetes Care* 2019;42:17–26.
- 30 Keser T, Gornik I, Vučković F, et al. Increased plasma N-glycome complexity is associated with higher risk of type 2 diabetes. *Diabetologia* 2017;60:2352–60.
- 31 Fisher P, Thomas-Oates J, Wood AJ, et al. The N-Glycosylation Processing Potential of the Mammalian Golgi Apparatus. *Front Cell Dev Biol* 2019;7:157.
- 32 Wittenbecher C, Štambuk T, Kuxhaus O, et al. Plasma N-Glycans as Emerging Biomarkers of Cardiometabolic Risk: A Prospective Investigation in the EPIC-Potsdam Cohort Study. *Diabetes Care* 2020;43:661–8.
- 33 Knežević A, Gornik O, Polašek O, et al. Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma N-glycans. *Glycobiology* 2010;20:959–69.
- 34 Nilsson A, Santoro A, Franceschi C, et al. Detrimental links between physical inactivity, metabolic risk and N-glycomic biomarkers of aging. *Exp Gerontol* 2019;124:110626.
- 35 Dong X, Mondello S, Kobeissy F, et al. LC–MS/Ms glycomics of idiopathic rapid eye movement sleep behavior disorder. *Electrophoresis* 2018;39:3096–103.
- 36 Kim T, Xie Y, Li Q, et al. Diet affects glycosylation of serum proteins in women at risk for cardiometabolic disease. *Eur J Nutr* 2021. doi:10.1007/s00394-021-02539-7. [Epub ahead of print: 26 Mar 2021].
- 37 Borodulin K, Tolonen H, Jousilahti P, et al. Cohort profile: the National FINRISK study. *Int J Epidemiol* 2018;47:696–696i.
- 38 Agakova A, Vučković F, Klarić L, et al. Automated integration of a UPLC glycomic profile. *Methods Mol Biol* 2017;1503:217–33.
- 39 Zhao S, Walsh I, Abrahams JL, et al. GlycoStore: a database of retention properties for glycan analysis. *Bioinformatics* 2018;34:3231–2.
- 40 Fruh SM. Obesity: risk factors, complications, and strategies for sustainable long-term weight management. *J Am Assoc Nurse Pract* 2017;29:S3–14.
- 41 Novokmet M, Lukić E, Vučković F, et al. Changes in IgG and total plasma protein glycomes in acute systemic inflammation. *Sci Rep* 2014;4:4347.
- 42 Freeman AM, Pennings N. Insulin Resistance. In: *StatPearls*. Treasure Island, FL: StatPearls Publishing, 2020. <http://www.ncbi.nlm.nih.gov/books/NBK507839/>
- 43 Warram JH, Martin BC, Krolewski AS, et al. Slow glucose removal rate and hyperinsulinemia precede the development of type II diabetes in the offspring of diabetic parents. *Ann Intern Med* 1990;113:909–15.
- 44 Clerc F, Reiding KR, Jansen BC, et al. Human plasma protein N-glycosylation. *Glycoconj J* 2016;33:309–43.
- 45 McMillan DE. Increased levels of acute-phase serum proteins in diabetes. *Metabolism* 1989;38:1042–6.
- 46 Alquoaq RS, Kasabri V, Naffa R, et al. Cross-sectional correlates of myeloperoxidase and alpha-1-acid glycoprotein with adiposity, atherogenic and hematological indices in metabolic syndrome patients with or without diabetes. *Ther Adv Endocrinol Metab* 2018;9:283–91.
- 47 Akbay E, Yetkin I, Ersoy R, et al. The relationship between levels of alpha-1-acid glycoprotein and metabolic parameters of diabetes mellitus. *Diabetes Nutr Metab* 2004;17:331–5.
- 48 Keser T, Tijardović M, Gornik I, et al. High-Throughput and site-specific N-glycosylation analysis of human alpha-1-acid glycoprotein offers a great potential for new biomarker discovery. *Mol Cell Proteomics* 2021:100044.
- 49 Hanley AJG, Williams K, Gonzalez C, et al. Prediction of type 2 diabetes using simple measures of insulin resistance: combined results from the San Antonio heart study, the Mexico City diabetes study, and the insulin resistance atherosclerosis study. *Diabetes* 2003;52:463–9.
- 50 Narayan KMV, Boyle JP, Thompson TJ, et al. Effect of BMI on lifetime risk for diabetes in the U.S. *Diabetes Care* 2007;30:1562–6.
- 51 Jo A, Mainous III AG. Informational value of percent body fat with body mass index for the risk of abnormal blood glucose: a nationally representative cross-sectional study. *BMJ Open* 2018;8:e019200.
- 52 Kuusela P, Saraswat M, Joensuu S, et al. Changes in plasma protein levels as an early indication of a bloodstream infection. *PLoS One* 2017;12:e0172987.
- 53 Prabantu VM, Naveenkumar N, Srinivasan N. Influence of disease-causing mutations on protein structural networks. *Front Mol Biosci* 2020;7:620554.
- 54 Gummesson A, Björnson E, Fagerberg L, et al. Longitudinal plasma protein profiling of newly diagnosed type 2 diabetes. *EBioMedicine* 2021;63:103147.
- 55 Tetik S, Kiliç A, Aksoy H, et al. Oxidative stress causes plasma protein modification. *Indian J Exp Biol* 2015;53:25–30.
- 56 Reily C, Stewart TJ, Renfrow MB, et al. Glycosylation in health and disease. *Nat Rev Nephrol* 2019;15:346–66.
- 57 Harris MI, Klein R, Welborn TA, et al. Onset of NIDDM occurs at least 4-7 yr before clinical diagnosis. *Diabetes Care* 1992;15:815–9.
- 58 Chatterjee S, Khunti K, Davies MJ. Type 2 diabetes. *The Lancet* 2017;389:2239–51.