



Plasma *N*-Glycans as Emerging Biomarkers of Cardiometabolic Risk: A Prospective Investigation in the EPIC-Potsdam Cohort Study

Diabetes Care 2020;43:661–668 | <https://doi.org/10.2337/dc19-1507>

Clemens Wittenbecher,^{1,2,3}
 Tamara Štambuk,⁴ Olga Kuxhaus,^{1,3}
 Najda Rudman,⁴ Frano Vučković,⁵
 Jerko Štambuk,⁵ Catarina Schiborn,^{1,3}
 Dario Rahelić,⁶ Stefan Dietrich,¹
 Olga Gornik,^{4,5} Markus Perola,^{7,8}
 Heiner Boeing,¹ Matthias B. Schulze,^{1,3,9}
 and Gordan Lauc^{4,5}

OBJECTIVE

Plasma protein *N*-glycan profiling integrates information on enzymatic protein glycosylation, which is a highly controlled ubiquitous posttranslational modification. Here we investigate the ability of the plasma *N*-glycome to predict incidence of type 2 diabetes and cardiovascular diseases (CVDs; i.e., myocardial infarction and stroke).

RESEARCH DESIGN AND METHODS

Based on the prospective European Prospective Investigation of Cancer (EPIC)-Potsdam cohort ($n = 27,548$), we constructed case-cohorts including a random subsample of 2,500 participants and all physician-verified incident cases of type 2 diabetes ($n = 820$; median follow-up time 6.5 years) and CVD ($n = 508$; median follow-up time 8.2 years). Information on the relative abundance of 39 *N*-glycan groups in baseline plasma samples was generated by chromatographic profiling. We selected predictive *N*-glycans for type 2 diabetes and CVD separately, based on cross-validated machine learning, nonlinear model building, and construction of weighted prediction scores. This workflow for CVD was applied separately in men and women.

RESULTS

The *N*-glycan-based type 2 diabetes score was strongly predictive for diabetes risk in an internal validation cohort (weighted C-index 0.83, 95% CI 0.78–0.88), and this finding was externally validated in the Finland Cardiovascular Risk Study (FINRISK) cohort. *N*-glycans were moderately predictive for CVD incidence (weighted C-indices 0.66, 95% CI 0.60–0.72, for men; 0.64, 95% CI 0.55–0.73, for women). Information on the selected *N*-glycans improved the accuracy of established and clinically applied risk prediction scores for type 2 diabetes and CVD.

CONCLUSIONS

Selected *N*-glycans improve type 2 diabetes and CVD prediction beyond established risk markers. Plasma protein *N*-glycan profiling may thus be useful for risk stratification in the context of precisely targeted primary prevention of cardiometabolic diseases.

¹Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany

²Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA

³German Center for Diabetes Research (DZD), München-Neuherberg, Germany

⁴Faculty of Pharmacy and Biochemistry, University of Zagreb, Zagreb, Croatia

⁵Genos Glycoscience Research Laboratory, Zagreb, Croatia

⁶University Clinics for Diabetes, Endocrinology and Metabolism, School of Medicine, University of Zagreb, Zagreb, Croatia

⁷National Institute for Health and Welfare, Helsinki, Finland

⁸Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Helsinki, Finland

⁹Institute of Nutritional Sciences, University of Potsdam, Nuthetal, Germany

Corresponding author: Matthias B. Schulze, mschulze@dife.de

Received 29 July 2019 and accepted 10 December 2019

This article contains Supplementary Data online at <https://care.diabetesjournals.org/lookup/suppl/doi:10.2337/dc19-1507/-/DC1>.

M.B.S. and G.L. share senior authorship.

© 2020 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. More information is available at <https://www.diabetesjournals.org/content/license>.

Cardiometabolic diseases, such as type 2 diabetes and cardiovascular disease (CVD; i.e., myocardial infarction, and stroke), constitute a major burden on public health systems. Primary prevention of cardiometabolic end points is therefore particularly advantageous to improve population health. Accurate risk prediction is an important tool to allocate resources for primary prevention to those who will maximally benefit.

Moreover, refined molecular phenotyping may differentiate between subgroups of high-risk individuals with distinct pathophysiological mechanisms of action. Thus, broadening the spectrum of molecular risk markers is a promising step toward precisely tailored prevention concepts. Specifically, it is required to discover biomarkers from biological domains that have thus far been underrepresented in risk prediction efforts.

N-Glycosylation denotes an essential and ubiquitous, complexly regulated, post-translational modification at defined asparagine residuals in target proteins. N-glycosylation enhances both structural and functional protein diversity (1,2) and is fundamentally different from nonenzymatic glycation, reflected through HbA_{1c} levels. Protein glycosylation is essential for self-recognition in biological systems, organization of intermolecular interactions, and determination of the spatial configuration of molecules (3). Plasma protein N-glycosylation patterns are partially genetically determined (4) and remarkably stable within healthy individuals (5). Composition of the plasma N-glycome can be reliably measured in chromatography-based high-throughput profiling approaches (6). Altogether, these facts suggest potential clinical utility of monitoring N-glycan patterns as early indicators of pathogenic processes.

Alterations in N-glycan profiles were cross-sectionally linked to various pathological conditions, such as obesity (7), inflammation (8), type 2 diabetes (9), and cardiovascular risk factors (10), indicating a potential link between N-glycan profiles and cardiometabolic risk. However, powerful prospective investigations that link N-glycomics to the onset of cardiometabolic end points are scarce.

Here we present the first human population study on the predictive value of the total plasma protein N-glycome for future incidence of major cardiometabolic end points. We developed plasma

N-glycan-based risk scores to predict type 2 diabetes and CVD in baseline-healthy participants of the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study and evaluated the predictive performance of these scores in comparison with established risk prediction tools.

RESEARCH DESIGN AND METHODS

Study Population

The prospective EPIC-Potsdam cohort study includes 27,548 participants (16,644 women and 10,904 men), who were recruited within an age range of 35–65 years from the general population between 1994 and 1998 (11). Baseline assessments comprised anthropometric measures and blood sampling by qualified medical personnel. Blood pressure was measured with the individual sitting with the arm elevated at heart level by using oscillometric devices (BOSO-Oscillomat; Bosch & Sohn, Jungingen, Germany), and the mean of the second and third reading was used. Lifestyle, sociodemographic characteristics, and current health status were assessed with validated questionnaires and in face-to-face interviews (12). Participants were then actively contacted by sending out questionnaires and, if necessary, by telephone every 2–3 years with response rates between 90 and 96% per follow-up round (13). All participants gave informed consent for biomedical research, and the study was approved by the Ethics Committee of the State of Brandenburg, Germany (11).

Nested case-cohorts were constructed for efficient studies into molecular phenotypes (Supplementary Fig. 1). From all participants who provided blood at baseline ($n = 26,437$), a random sample (subcohort, $n = 2,500$) was drawn. In addition, all incident diabetes and CVD cases that occurred in the full cohort until a specified censoring date were included. Statistically accounting for the oversampling of cases, this design provides unbiased risk estimates for the full cohort.

For type 2 diabetes, the censoring date was 31 August 2005 (820 incident cases). After exclusion of participants with missing follow-up information, prevalent diabetes at recruitment, insufficient blood specimens, or nonverifiable information on diabetes incidence, the analytical sample comprised 2,813 participants (1,578

women and 1,235 men), including 743 participants with incident type 2 diabetes, from whom 64 were randomly included in the subcohort. The median follow-up time for type 2 diabetes was 6.5 years (interquartile range 6–8.6 years). This study sample was split into a training set (two-thirds of the population) and a validation set (one-third of the population) applying random selection within strata according to sex and case status.

For CVD, 508 incident primary cardiovascular events occurred until censoring date (31 November 2006). After equivalent exclusions (using prevalent and nonverifiable CVD instead of diabetes as the exclusion criterion), the CVD sample comprised 2,558 participants (1,481 women and 1,077 men), including 418 participants with incident CVD (44 randomly included in the subcohort). The median follow-up time for CVD was 8.3 years (interquartile range 7.5–9.2 years).

Case Ascertainment

Hints of potential incidence of diabetes and CVD were detected based on self-reports of disease diagnosis, of disease-related medication or dietary treatments in the regular follow-up procedure, or based on information from other health-related data (such as death certificates for CVD). For all participants with a potential incident end point, the diagnosis was verified by contacting the treating primary care physician. Only participants with physician-verified diagnosis of type 2 diabetes (ICD-10 code: E11), myocardial infarction or stroke (ICD-10 codes: I21 for acute myocardial infarction, I63.0–I63.9 for ischemic stroke, I61.0–I61.9 for intracerebral hemorrhage, I60.0–I60.9 for subarachnoid hemorrhage, and I64.0–I64.9 for unspecified stroke) and a diagnosis date after the baseline examination were considered incident cases. For CVD, the World Health Organization MONICA (MONItoring Trends and Determinants on Cardiovascular Diseases) criteria were applied to classify events into possible, probable, and definite (for further details see Supplementary Note 1).

Laboratory Measurements

Individual plasma samples (10 μ L) were denatured by incubation with SDS (Invitrogen) at 65°C for 10 min. N-Glycans were then released with the addition of 1.2 units of peptide:N-glycosidase F (Promega) and overnight incubation at 37°C (14).

The released *N*-glycans were labeled with 2-aminobenzamide (Sigma-Aldrich), separated by hydrophilic interaction liquid chromatography on an Acquity Ultra-Performance Liquid Chromatography (UPLC Technology) instrument (Waters), and quantified with a fluorescence detector. Data were processed using an automated integration method (15), and chromatograms were separated into 39 glycan peaks (GPs), where the integrated area under each peak relative to the total integrated area was used as a measure of relative abundance of the designated *N*-glycan group (Supplementary Note 2 and Supplementary Fig. 2). Supplementary Table 1 matches the detected peaks to the most abundant glycan structures. Samples were randomly distributed across batches, and laboratory personnel were blinded to case status. Total cholesterol, HDL cholesterol, random glucose, and HbA_{1c} were measured using an automatic ADIVA 1650 analyzer (Siemens Medical Solutions, Erlangen, Germany) at the University Clinic Tübingen, Tübingen, Germany.

Statistical Analysis

First, *N*-glycan data were adjusted for age by applying multiple fractional polynomial (MFP) regression separately in men and women. The age-adjusted residuals were used to construct an *N*-glycan score to predict type 2 diabetes by applying the following steps (Supplementary Fig. 3 and Supplementary Note 3):

1. Machine learning–based preselection: *N*-glycan type 2 diabetes risk markers were preselected by applying random survival forest (RSF) in combination with a backward selection procedure in the training set (16).
2. Building the model: MFP testing in Cox proportional hazards models was used to omit redundant information and select the optimal power transformation for nonlinear associations (17).
3. Deriving a weighted score: All GPs were variance standardized ($\mu = 0$, $SD = 1$). Betas from the Cox model from step 2 were used as weights to build the glycan-based type 2 diabetes prediction score (GS_{T2D}).

Type 2 diabetes prediction with the GS_{T2D} was evaluated in the independent validation cohort. Several models were

assessed for their ability to predict type 2 diabetes incidence:

1. an *N*-glycan score based on age-adjusted, type 2 diabetes–related *N*-glycans (GS_{T2D})
2. a weighted combination of the GS_{T2D} and age ($GS_{T2D} + \text{age}$)
3. the noninvasive German Diabetes Risk Score (GDRS) (18,19)
4. plasma HbA_{1c} and random glucose concentrations ($Gluc_{\text{Markers}}$)
5. a weighted combination of the GS_{T2D} and GDRS ($GS_{T2D} + \text{GDRS}$)
6. a weighted combination of the GS_{T2D} and $Gluc_{\text{Markers}}$ ($GS_{T2D} + Gluc_{\text{Markers}}$)

The GDRS relies on information on age, sex, anthropometry (waist circumference and height), lifestyle (physical activity, smoking status, and consumption of coffee, red meat, and whole grain), hypertension, and family history of diabetes.

The CVD analyses were stratified by sex and did not include an internal validation step because of power considerations. Based on the hypothesis of shared risk factors between diabetes and CVD, we first tested whether the GS_{T2D} was also linked to cardiovascular risk. In parallel, an exploratory selection of potential CVD-specific predictors was conducted analogous to the selection of diabetes-related *N*-glycans, with minor modifications (Supplementary Fig. 4):

1. Machine learning–based preselection: RSF selection was applied in men and women separately.
2. Building the model: MFP was used to combine exploratory selected *N*-glycans from step 1 with the GS_{T2D} (only included if significantly associated with CVD risk) allowing for nonlinearity.
3. Deriving a weighted score: Combination of predictors into a weighted score was only applied if more than one predictor was selected.

The recalibrated version of the American Heart Association (AHA) score (20) was used as an established prediction tool for comparison. The AHA score includes information on age, total and HDL cholesterol, systolic blood pressure, antihypertensive medication, smoking, and prevalent type 2 diabetes and relies on sex-specific risk equations.

To evaluate model predictions, the receiver operating characteristics (ROC) curve was used to visualize adequacy of model predictions, weighted C-indices (accounting for the case-cohort design) were estimated to evaluate the model discrimination between cases and non-cases, and the weighted continuous net reclassification index (NRI) was estimated to evaluate whether additional information on the *N*-glycans improved disease prediction with established risk prediction tools (21).

Statistical analyses were performed in SAS 9.4 and R 3.3.2 software. For RSFs we used the randomForestSRC package in R (<https://cran.r-project.org/web/packages/randomForestSRC/index.html>). For MFP modeling we used a modified version of the SAS macro by Sauerbrei et al. (<https://mfp.imbi.uni-freiburg.de/software>). For the estimation of weighted C-indices and weighted NRIs, we used SAS macros by Cook et al. (<https://ncook.bwh.harvard.edu/sas-macros.html>).

RESULTS

Descriptive Statistics

The distribution of baseline characteristics in EPIC-Potsdam participants is provided in Table 1. Among participants with incident cardiometabolic end points, median age, waist circumference, and total cholesterol were higher, while median height, HDL cholesterol plasma concentrations, and whole grain consumption were lower compared with the cohort-representative subcohort. Hypertension was more prevalent among participants with incident cardiometabolic diseases. Current smoking at baseline was most prevalent in participants with incident CVD in the study course. The distribution of total plasma protein *N*-glycans at baseline in the EPIC-Potsdam cohort and age-adjusted and sex-stratified intercorrelation of the assessed *N*-glycans are shown across participants with and without incident cardiometabolic diseases in Supplementary Table 2 and Supplementary Figs. 5–7. Sex-stratified and age-adjusted correlations of GPs with blood lipids (total and HDL cholesterol and triglycerides) are reported in Supplementary Table 3.

Glycan-Based Prediction of Type 2 Diabetes Incidence

Stepwise RSF selection in the training set indicated that 8 of 39 *N*-glycans were

Table 1—Distribution of baseline characteristics

	Subcohort*	Incident T2D	Incident CVD
Men			
Observations (n)	728	439	269
Age at recruitment (years)	51.1 (44.0, 57.9)	56.3 (49.5, 60.8)	57.9 (52.6, 62.1)
Height (cm)	175 (171, 179)	173 (169, 178)	173 (168, 178)
Waist circumference (cm)	92 (87, 99)	103 (97, 110)	96.5 (90, 105)
Sport, biking, gardening (h/week)	5.5 (2.5, 9.5)	5.0 (2.0, 9.5)	6.0 (2.5, 12)
Whole grain (50 g portions/day)	0.47 (0.11, 1.31)	0.30 (0.05, 1.02)	0.30 (0.04, 0.99)
Coffee (150 mL cups/day)	3 (1, 4)	2 (1, 4)	2 (2, 4)
Red meat (150 g portions/day)	0.32 (0.22, 0.47)	0.34 (0.22, 0.52)	0.33 (0.22, 0.47)
Total cholesterol (mg/dL)	210 (182, 236)	213 (187, 239)	214 (191, 239)
HDL cholesterol (mg/dL)	49.5 (41.9, 56.8)	42.9 (37.3, 48.9)	44.9 (38.9, 52)
Triglyceride (mg/dL)	133 (92.7, 197)	190 (139, 278)	147 (99.8, 242)
HbA _{1c} (%)	5.43 (5.12, 5.75)	6.12 (5.71, 6.78)	5.7 (5.43, 6.13)
HbA _{1c} (mmol/mol)	35.8 (32.5, 39.3)	43.4 (38.9, 50.6)	38.8 (35.8, 43.4)
Smoking categories (%)			
Never smoker	30	19	21
Former smoker	44	55	38
Smoker <20 units/day	16	14	23
Smoker ≥20 units/day	9	13	18
Prevalent hypertension (%)	54	83	75
Intake of antilipidemic drugs (%)	5	10	7
Intake of antihypertensive drugs (%)	15	37	31
Intake of acetylsalicylic acid (%)	6	13	10
Women			
Observations (n)	1,251	304	149
Age at recruitment	47.0 (40.1, 56.3)	57.0 (50.3, 61.1)	57.2 (49.5, 61.3)
Height (cm)	163 (159, 167)	162 (158, 166)	162 (159, 166)
Waist circumference (cm)	78 (71, 86)	94 (86, 102)	84 (76, 92)
Sport, biking, gardening (h/week)	4.0 (2.0, 7.5)	4.0 (1.0, 8.2)	3.5 (1.0, 7.0)
Whole grain (50 g portions/day)	0.77 (0.24, 1.66)	0.66 (0.23, 1.45)	0.69 (0.19, 1.54)
Coffee (150 mL cups/day)	2 (2, 4)	2 (2, 4)	3 (2, 4)
Red meat (150 g portions/day)	0.21 (0.13, 0.30)	0.23 (0.15, 0.33)	0.21 (0.11, 0.32)
Total cholesterol (mg/dL)	198 (173, 226)	213 (188, 243)	221 (189, 252)
HDL cholesterol (mg/dL)	59.1 (49.8, 69.3)	50.6 (43.0, 58.1)	57.9 (46.5, 68.7)
Triglyceride (mg/dL)	91.8 (68.6, 130)	153 (112, 205)	121 (80.8, 187)
HbA _{1c} (%)	5.33 (5.09, 5.67)	6.13 (5.68, 6.62)	5.55 (5.27, 5.99)
HbA _{1c} (mmol/mol)	34.8 (32.1, 38.5)	43.4 (38.6, 48.9)	37.2 (34.1, 42)
Smoking categories (%)			
Never smoker	58	58	57
Former smoker	25	29	20
Smoker <20 units/day	14	9	16
Smoker ≥20 units/day	3	4	7
Prevalent hypertension (%)	39	73	69
Intake of antilipidemic drugs (%)	3	10	6
Intake of antihypertensive drugs (%)	15	42	41
Intake of acetylsalicylic acid (%)	9	12	11

Data are median (interquartile range) unless otherwise indicated. The EPIC-Potsdam cohort is a population-based sample from the Potsdam region, Brandenburg, Germany. All of the participants included in this study are whites of European descent. *For this table, all participants with prevalent or incident type 2 diabetes or CVD were excluded from the random subcohort.

informative with regard to time to type 2 diabetes incidence, namely, GP10, GP18, GP20, GP22, GP26, GP32, GP35, and GP36 (Supplementary Fig. 8). MFP modeling included GP20, GP26, and GP36 as linear predictors and GP18, GP32, and GP35 as nonlinear terms ($\frac{1}{GP18}$, $GP32^2$, and $\frac{1}{\sqrt{GP35}}$), while GP10 and GP22 were omitted. After influential observations were excluded, this model was used to estimate standardized betas yielding the

following N-glycan score equation for type 2 diabetes prediction:

$$GS_{T2D} = 1.82 * \frac{1}{GP18} + 1.01 * GP20 + 1.97 * GP26 - 0.59 * GP32^2 - 1.47 * \frac{1}{\sqrt{GP35}} - 1.17 * GP36$$

In both the training and the validation cohort, GP36 was related to lower relative

risk of type 2 diabetes, whereas GP20 and GP26 were linked to higher risk. For GP18 and GP35, the power transformation reversed the order of values, so that the positive and negative risk estimate for the transformed terms translate into lower GP18-related and higher GP35-related type 2 diabetes risk with higher values on the original scale. GP32 was related to lower risk in the training cohort but was noninformative for type 2 diabetes risk in the validation cohort. The type 2 diabetes hazard ratio (HR) for 1 SD higher GS_{T2D} in the validation cohort was 3.73 (95% CI 3.0–4.63) (Table 2).

ROC curves of the GS_{T2D} and $GS_{T2D} + age$ are displayed in Fig. 1, in comparison with the GDRS (Fig. 1A) and established clinical type 2 diabetes risk markers (HbA_{1c} and blood glucose concentrations) (Fig. 1B) in the validation cohort. The C-index of 0.83 (95% CI 0.78–0.88) for type 2 diabetes prediction with GS_{T2D} in the validation cohort indicated strong discriminatory ability of the selected age-adjusted plasma N-glycans alone. Addition of age ($GS_{T2D} + age$) improved the N-glycan-based type 2 diabetes prediction (C-index 0.87, 95% CI 0.82–0.91), which was comparable to the predictive performance of the GDRS (C-index 0.88, 95% CI 0.82–0.92). A weighted combination of the two scores ($GS_{T2D} + GDRS$) conferred better discrimination, as indicated by a C-index of 0.9 (95% CI 0.85–0.94). Improvement of the model predictions by $GS_{T2D} + GDRS$ compared with the reference model (GDRS) was corroborated by a positive NRI of 0.62 (95% CI 0.53–0.71). Comparison of the reclassification indices among participants with incident type 2 diabetes (RI_{cases} 0.42, 95% CI 0.37–0.46) and without type 2 diabetes incidence ($RI_{noncases}$ 0.2, 95% CI 0.19–0.22) clarified that the net reclassification improvement was predominantly driven by assignment of higher risk to cases (i.e., improved model sensitivity) (Table 3).

Similarly, the combination of GS_{T2D} with HbA_{1c} and glucose (C-index 0.91, 95% CI 0.87–0.95) outperformed type 2 diabetes prediction with these established clinical risk markers alone (C-index 0.88, 95% CI 0.83–0.92). Again, net reclassification analyses underpinned the added predictive value of glycans compared with HbA_{1c} + glucose alone (NRI 0.87, 95% CI 0.79–0.94) and clarified that this model improvement was dominated by higher sensitivity (i.e., assignment of

Table 2—Plasma N-glycans and risk of type 2 diabetes and CVD

Type 2 diabetes			
Training cohort		Validation cohort	
Plasma N-glycan	HR per SD (95% CI)	Plasma N-glycan	HR per SD (95% CI)
GP18 ⁻¹	2.37 (1.96–2.86)	GP18 ⁻¹	1.95 (1.47–2.58)
GP20	1.58 (1.33–1.88)	GP20	1.55 (1.22–1.97)
GP26	2.57 (2.13–3.09)	GP26	1.91 (1.49–2.43)
GP32 ²	0.79 (0.64–0.98)	GP32 ²	1.05 (0.74–1.48)
GP35 ^{-0.5}	0.50 (0.42–0.58)	GP35 ^{-0.5}	0.63 (0.50–0.80)
GP36	0.56 (0.48–0.66)	GP36	0.53 (0.42–0.66)
GS _{T2D} *	4.20 (3.57–4.94)	GS _{T2D} *	3.73 (3.00–4.63)
CVD			
Men		Women	
Plasma N-glycan	HR per SD (95% CI)	Plasma N-glycan	HR per SD (95% CI)
GS _{T2D}	1.23 (1.09–1.40)		
GP23	1.55 (1.29–1.87)		
GP16	0.50 (0.40–0.63)		
GS _{CVD} †	1.89 (1.57–2.27)	GP5‡	0.70 (0.57–0.86)

Individual N-glycans were mutually adjusted and controlled for age. *GS_{T2D}, N-glycan score for type 2 diabetes prediction, a weighted combination of the selected N-glycans. Weights based on risk estimates in the training cohort ($n_{\text{subcohort}} = 1,417$, $n_{\text{cases}} = 495$, overlap 44) and evaluated in the validation cohort ($n_{\text{subcohort}} = 717$, $n_{\text{cases}} = 248$, overlap 20). †GS_{CVD}, the CVD prediction score in men was derived as a weighted combination of GS_{T2D}, GP23, and GP16 (men: $n_{\text{subcohort}} = 841$, $n_{\text{cases}} = 269$, overlap = 33). ‡In women, only GP5 was selected as robust risk marker, hence no score was derived (women: $n_{\text{subcohort}} = 1,343$, $n_{\text{cases}} = 149$, overlap = 11).

higher risk to participants with incident type 2 diabetes) (Table 3).

Glycan-Based Prediction of Incident Cardiovascular Events

We first tested whether the GS_{T2D} was also linked to cardiovascular risk in sex-stratified analyses in the case-cohort. In men, higher GS_{T2D} was associated significantly with higher risk of cardiovascular events. In addition, the exploratory selection procedure suggested CVD predictive ability of GP16, GP23, and GP29 in men (Supplementary Fig. 9). The association of GP29 was rendered nonsignificant when additionally adjusted for GS_{T2D}. Therefore, the N-glycan-based CVD score for men (GS_{CVD}) was constructed as a weighted combination of GP16 (associated with lower CVD risk) and GP23 and GS_{T2D} (both associated with higher CVD risk). Higher GS_{CVD} points were associated with a higher risk of cardiovascular events in men in an age-adjusted model (HR per SD 1.89, 95% CI 1.57–2.27) (Table 2). The weighted C-index for GS_{CVD} in men of 0.66 (95% CI 0.6–0.72) suggested moderate discrimination of age-adjusted N-glycans alone. The C-index of 0.71 (95% CI 0.65–0.77) for GS_{CVD}+age corresponds to good discrimination (Fig. 1C). However, discrimination by the AHA score

was slightly better (C-index 0.73, 95% CI 0.67–0.78), and the combination of GS_{CVD}+AHA score did not yield a higher C-index (0.72, 95% CI 0.66–0.78). Still, closer inspection of model improvement (GS_{CVD}+AHA score vs. AHA score) with the NRI (0.31, 95% CI 0.24–0.39) suggested improved prediction due to increased specificity (i.e., assignment of lower risk to participants without incident cardiovascular events) (Table 3).

In women, GS_{T2D} was not independently associated with CVD risk, and exploratory selection pointed at GP5 as a CVD risk marker (Supplementary Fig. 10). GP5 was inversely associated with incident cardiovascular events in women (HR per SD 0.7, 95% CI 0.57–0.86). The C-index for GP5 (0.64, 95% CI 0.55–0.73) indicated moderate predictive ability of this N-glycan alone, while good model predictions were generated for GP5+age (C-index 0.75, 95% CI 0.66–0.82) (Fig. 1D). Discrimination by the AHA score (C-index 0.74, 95% CI 0.65–0.82) was comparable to GP5+age. A combination of GP5 and the AHA score in women yielded a higher C-index of 0.77 (95% CI 0.69–0.85). The NRI of 0.58 (95% CI 0.49–0.67) corroborated net model improvement by combining GP5 with the AHA score. Higher sensitivity was the main driver of the

improved predictions (RI_{cases} 0.5, 95% CI 0.46–0.54) (Table 3).

External Validation and Sensitivity Analyses

We validated the association with type 2 diabetes risk in a case-control study nested within the Finland Cardiovascular Risk Study (FINRISK) cohort, including 37 participants with incident type 2 diabetes and 37 age- and sex-matched participants who remained diabetes free and served as control subjects (study sample was described elsewhere [22]). In this sample, 1 SD higher GS_{T2D} was associated with a doubling of type 2 diabetes risk (odds ratio 2.03, 95% CI 1.13–3.63), and the C-index of 0.76 (95% CI 0.65–0.87) indicated strong type 2 diabetes prediction over 10 years (Supplementary Fig. 11).

Sensitivity analyses demonstrated that adjustment for lipid lowering and antihypertensive medication did not substantially alter the association of glycan scores with cardiometabolic risk. The CVD results were robust against exclusion of participants with prevalent diabetes at baseline and after restriction to participants with a definite cardiovascular event according to the MONICA criteria. Associations of GS_{CVD} (men) and GP5 (women) did not substantially differ between stroke and myocardial infarction (Supplementary Table 4).

CONCLUSIONS

From a panel of 39 N-glycan groups originating from total plasma proteins, 6 N-glycan residuals exhibited strong, robust, and mutual independent associations with type 2 diabetes risk. A score based on these N-glycans was strongly predictive for type 2 diabetes incidence and improved risk prediction with known risk factors. The strong link between the glycan score and type 2 diabetes risk was replicated in internal and external validation studies. Our results further suggested that information on age-adjusted N-glycans alone was moderately predictive for incident CVD and that additional information on N-glycans may improve CVD prediction with the established clinical AHA score. In summary, we provide first-line evidence that plasma N-glycome screening generates valuable markers for the risk of developing type 2 diabetes and CVD.

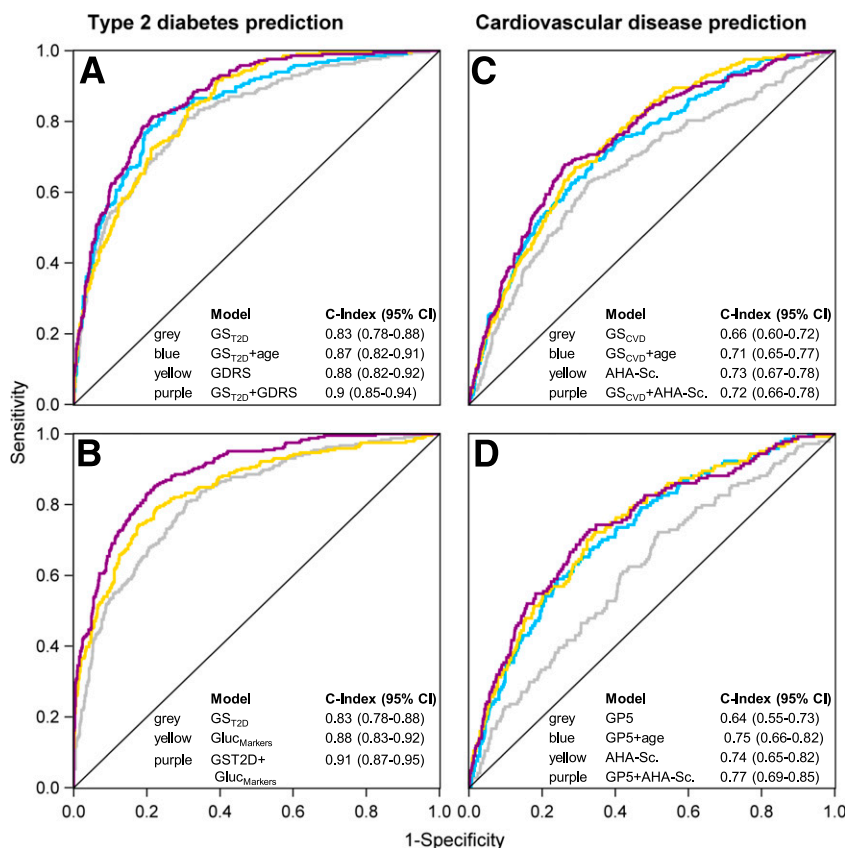


Figure 1—N-glycan-based prediction of cardiometabolic diseases. Total plasma protein N-glycomics-based prediction of type 2 diabetes compared with the noninvasive GDRS (A) and clinical markers of glucose metabolism (B), and CVD prediction compared with the AHA risk score (Sc.) in men (C) and in women (D).

The cross-sectional link between plasma protein glycosylation patterns and type 2 diabetes and cardiometabolic risk markers is well documented (8,9,22–24), and a small case-control study reported a link between baseline N-glycan profiles and type 2 diabetes risk (22). Our internally validated findings now demonstrate that a plasma N-glycan score, GS_{T2D}, accurately predicts type 2 diabetes incidence. Moreover, we externally validated these results in the independent FINRISK study.

Furthermore, we showed that the GS_{T2D} improved prediction with the widely applied GDRS (25) and with plasma HbA_{1c} and random blood glucose concentration. The probable source proteins of the type 2 diabetes-related N-glycans comprise (26) α1-acid glycoprotein (branched, highly sialylated GP26, GP32, GP35, and GP36) and transferrin, hemopexin, and α-antitrypsin (GP18 and GP20). Through transport, signaling, and intermolecular interactions, these proteins are involved in different aspects of metabolism and immunity (26)

and were previously associated with increased risk of several diseases (27), including type 2 diabetes (28–30) and CVD (31–33). Importantly, altered glycosylation of plasma proteins affects their function. A recent study, for example, showed that incomplete glycosylation of lipoproteins leads to accumulation of cholesterol in cells (34). Prospective human studies with integrated glycomics and glycoproteomics data will be able to address the interaction between protein concentration and glycosylation patterns in relation to disease risk and are thus highly warranted.

In sex-stratified analyses, age-adjusted plasma N-glycans were moderately predictive for CVD incidence. Moreover, N-glycans improved CVD prediction with the AHA score, which combines age with the most established clinical risk factors for CVD. GP5 particularly assigned higher risk to women with incident CVD. Among men, the net reclassification improvement was determined by assigning lower risk to participants without incident cardiovascular events, which may be less relevant from a clinical standpoint. The probable source proteins of exclusively CVD-related GPs are IgG (GP5 and GP16) and IgA (GP23). Hence, type 2 diabetes-related markers were also useful for CVD prediction in men, which may be interpreted as a hint toward pathways related to both end points. The fact that all CVD-specific glycan markers stemmed from Ig’s suggests a particular role of glycosylation-dependent immune response in CVD etiology.

Several studies suggest that protein N-glycosylation is mechanistically involved in type 2 diabetes etiology. For instance, the *ST6GAL1* gene has been linked to type 2 diabetes risk, where its risk variant was related to increased expression of the encoded α2,6-sialyltransferase in β-cells and to decreased stimulated insulin secretion (35). It was also recently shown in mouse models that improper sialylation of IgG can cause diabetes (36). Other candidate pathways that may link the N-glycome to cardiometabolic diseases comprise glycosylation-dependent protein trafficking (e.g., surface distribution of GLUTs) (37,38), the hexosamine pathway (39), and adaptive immunity (40). N-glycan-based biomarkers may thus reflect the activity of intracellular pathways that predetermine diabetes onset, and integrated molecular epidemiological

Table 3—Analysis of prediction improvement with established prediction tools through additional information on risk-related N-glycans

	R _{cases} (95% CI)	R _{noncases} (95% CI)	NRI (95% CI)
Type 2 diabetes risk			
GDRS vs. GDRS+GS _{T2D}	0.42 (0.37–0.46)	0.20 (0.19–0.22)	0.62 (0.53–0.71)
Gluc _{markers} vs. GDRS+GS _{T2D}	0.64 (0.61–0.68)	0.22 (0.21–0.23)	0.87 (0.79–0.94)
CVD risk men			
AHA score vs. AHA score+GS _{CVD}	0.01 (–0.03 to 0.04)	0.31 (0.30–0.32)	0.31 (0.24–0.39)
CVD risk women			
AHA score vs. AHA score+GP5	0.50 (0.46–0.54)	0.08 (0.07–0.09)	0.58 (0.49–0.67)

and experimental studies to elucidate these pathways are highly encouraged by our observations.

To our knowledge, this is the largest prospective study linking *N*-glycomics to cardiometabolic end points. Limitations of this study are related to sample size and validation of glycan-based CVD prediction, to the limited ability to test robustness of glycan-based disease prediction across different ethnicities, and to the laboratory measurement approach, and will be discussed in the following.

The link between GS_{T2D} and type 2 diabetes risk in the external FINRISK cohort was strong, but the C-index was slightly lower compared with EPIC-Potsdam analyses. However, the matched case-control design, which is not optimal to generate generalizable prediction estimates, and some imprecision due to the relatively small sample size in FINRISK, may explain these minor differences. Overall, our results support the generalizable value of plasma *N*-glycans for type 2 diabetes prediction in populations of middle-European descent. However, validation of our findings in other races and ethnicities is warranted.

The power to detect CVD-related *N*-glycans was lower due to the sex-stratified analysis design and slightly fewer incident cases, and we therefore renounced the sample split. Cross-validated RSF should still have selected robust predictors. Nonetheless, external validation of our CVD results is warranted.

It is noteworthy that some of the *N*-glycans were highly correlated so that the selection of other, closely related *N*-glycans as risk markers may have resulted in similarly accurate predictions. Moreover, we explored *N*-glycan residuals from total plasma proteins for cardiometabolic risk markers, which implies that the interpretation of our findings remains somewhat speculative with regard to the source proteins. Our observations, for example, suggest a targeted investigation of Ig-specific *N*-glycosylation patterns in relation to CVD risk.

In this work we have demonstrated that profiling of the total plasma protein *N*-glycome generates highly informative predictors of the future onset of cardiometabolic end points. The *N*-glycan-based prediction of type 2 diabetes incidence was stronger, and *N*-glycan predictors for cardiovascular end points only partially overlapped with those related to type 2

diabetes. Moreover, our results suggest that *N*-glycan biomarkers improve prediction of type 2 diabetes and CVD when combined with established risk scores. From a biological perspective, specific *N*-glycan structures likely reflect distinct disease-related pathways. We therefore conclude that plasma *N*-glycan biomarkers should complement the toolbox of novel biomarkers that are considered for risk stratification in precision prevention approaches.

Acknowledgments. The authors thank the Human Study Centre (HSC) of the German Institute of Human Nutrition Potsdam-Rehbrücke, namely, the trustee and the data hub, for data processing, the biobank for the processing of biological samples, and Manuela Bergmann, the head of the HSC, for contribution to the study design and leading the underlying processes of data generation. Furthermore, the authors thank all EPIC-Potsdam participants for their invaluable contribution to the study.

Funding. The work was supported by the Federal Ministry of Science, Germany (grant no. 01 EA 9401) and the European Union (grant no. SOC 95201408 05 F02) for the recruitment phase of the EPIC-Potsdam study, by the German Cancer Aid (grant no. 70-2488-Ha I) and the European Community (grant no. SOC98200769 05 F02) for the follow-up of the EPIC-Potsdam study, and by a grant from the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung) to the German Center for Diabetes Research (DZD, grant 82DZD00302). Plasma glycome analysis was performed in the Genos Glycoscience Research Laboratory and partly supported by the European Union's Horizon 2020 Framework Programme grants IMforFUTURE (grant agreement no. 721815) and GlySign (grant agreement no. 722095) as well as by the European Structural and Investment Funds IRI grant (no. KK.01.2.1.01.0003), CEKOM grant (no. KK.01.2.2.03.0006), and Croatian National Centre of Research Excellence in Personalized Healthcare grant (no. KK.01.1.1.01.0010).

Duality of Interest. G.L. is the founder and chief executive officer of Genos Ltd, a private research organization that specializes in high-throughput glycomics analysis and has several patents in this field. F.V., J.Š., and O.G. are employees of Genos Ltd. No other potential conflicts of interest relevant to this article were reported.

Author Contributions. C.W., O.G., M.B.S., and G.L. designed the study. T.Š., N.R., F.V., J.Š., and D.R. were involved in laboratory measurements and processing of the glycomics data. C.S. and S.D. contributed to prediction methods. C.W. and O.K. conducted the statistical analysis. C.W. drafted the manuscript. C.W., M.B.S., and G.L. are responsible for the content of the manuscript. All authors contributed to the interpretation of the results and critically revised the manuscript. C.W. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. Lauc G, Pezer M, Rudan I, Campbell H. Mechanisms of disease: the human *N*-glycome. *Biochim Biophys Acta* 2016;1860:1574–1582
2. Gornik O, Pavić T, Lauc G. Alternative glycosylation modulates function of IgG and other proteins - implications on evolution and disease. *Biochim Biophys Acta* 2012;1820:1318–1326
3. Varki A. Biological roles of glycans. *Glycobiology* 2017;27:3–49
4. Krištić J, Zaytseva OO, Ram R, et al. Profiling and genetic control of the murine immunoglobulin G glycome. *Nat Chem Biol* 2018;14:516–524
5. Gornik O, Wagner J, Pucić M, Knezević A, Redžić I, Lauc G. Stability of *N*-glycan profiles in human plasma. *Glycobiology* 2009;19:1547–1553
6. Adamczyk B, Stöckmann H, O'Flaherty R, Karlsson NG, Rudd PM. High-throughput analysis of the plasma *N*-glycome by UHPLC. In *High-Throughput Glycomics and Glycoproteomics*. Lauc G, Wuhler M, Eds. New York, NY, Humana Press, 2017, p. 97–108
7. Russell AC, Kepka A, Trbojević-Akmačić I, et al. Increased central adiposity is associated with pro-inflammatory immunoglobulin G *N*-glycans. *Immunobiology* 2019;224:110–115
8. 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
9. Lemmers RFH, Vilaj M, Urda D, et al. IgG glycan patterns are associated with type 2 diabetes in independent European populations. *Biochim Biophys Acta, Gen Subj* 2017;1861:2240–2249
10. Menni C, Gudelj I, Macdonald-Dunlop E, et al. Glycosylation profile of immunoglobulin G is cross-sectionally associated with cardiovascular disease risk score and subclinical atherosclerosis in two independent cohorts. *Circ Res* 2018;122:1555–1564
11. Boeing H, Korfmann A, Bergmann MM. Recruitment procedures of EPIC-Germany. *European Investigation into Cancer and Nutrition. Ann Nutr Metab* 1999;43:205–215
12. Boeing H, Wahrendorf J, Becker N. EPIC-Germany—a source for studies into diet and risk of chronic diseases. *European Investigation into Cancer and Nutrition. Ann Nutr Metab* 1999;43:195–204
13. Bergmann MM, Bussas U, Boeing H. Follow-up procedures in EPIC-Germany—data quality aspects. *European Prospective Investigation into Cancer and Nutrition. Ann Nutr Metab* 1999;43:225–234
14. Pavić T, Dilber D, Kifer D, et al. *N*-glycosylation patterns of plasma proteins and immunoglobulin G in chronic obstructive pulmonary disease. *J Transl Med* 2018;16:323
15. Agakova A, Vučković F, Klarić L, Lauc G, Agakov F. Automated integration of a UPLC glycomic profile. *Methods Mol Biol* 2017;1503:217–233
16. Dietrich S, Floegel A, Troll M, et al. Random survival forest in practice: a method for modelling complex metabolomics data in time to event analysis. *Int J Epidemiol* 2016;45:1406–1420
17. Sauerbrei W, Meier-Hirmer C, Benner A, Royston P. Multivariable regression model

- building by using fractional polynomials: description of SAS, STATA and R programs. *Comput Stat Data Anal* 2006;50:3464–3485
18. Mühlenbruch K, Ludwig T, Jeppesen C, et al. Update of the German Diabetes Risk Score and external validation in the German MONICA/KORA study. *Diabetes Res Clin Pract* 2014;104:459–466
19. Schulze MB, Hoffmann K, Boeing H, et al. An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. *Diabetes Care* 2007;30:510–515
20. de las Heras Gala T, Geisel MH, Peters A, et al.; KORA Study Group and the Heinz Nixdorf Recall Study Group. Recalibration of the ACC/AHA risk score in two population-based German cohorts. *PLoS One* 2016;11:e0164688
21. Cook NR, Ridker PM. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. *Ann Intern Med* 2009;150:795–802
22. Keser T, Gornik I, Vučković F, et al. Increased plasma N-glycome complexity is associated with higher risk of type 2 diabetes. *Diabetologia* 2017;60:2352–2360
23. Adua E, Memarian E, Russell A, et al. High throughput profiling of whole plasma N-glycans in type II diabetes mellitus patients and healthy individuals: a perspective from a Ghanaian population. *Arch Biochem Biophys* 2019;661:10–21
24. 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;20:959–969
25. Paprott R, Mühlenbruch K, Mensink GB, et al. Validation of the German Diabetes Risk Score among the general adult population: findings from the German Health Interview and Examination Surveys. *BMJ Open Diabetes Res Care* 2016;4:e000280
26. Clerc F, Reiding KR, Jansen BC, Kammeijer GSM, Bondt A, Wührer M. Human plasma protein N-glycosylation. *Glycoconj J* 2016;33:309–343
27. Fischer K, Kettunen J, Würtz P, et al. Biomarker profiling by nuclear magnetic resonance spectroscopy for the prediction of all-cause mortality: an observational study of 17,345 persons. *PLoS Med* 2014;11:e1001606
28. Schmidt MI, Duncan BB, Sharrett AR, et al. Markers of inflammation and prediction of diabetes mellitus in adults (Atherosclerosis Risk in Communities study): a cohort study. *Lancet* 1999;353:1649–1652
29. Higai K, Azuma Y, Aoki Y, Matsumoto K. Altered glycosylation of α 1-acid glycoprotein in patients with inflammation and diabetes mellitus. *Clin Chim Acta* 2003;329:117–125
30. Zacharski LR, Shamayeva G, Chow BK, DePalma RG. Ferritin and percent transferrin saturation levels predict type 2 diabetes risk and cardiovascular disease outcomes. *Curr Diabetes Rev* 2017;13:428–436
31. Yin X, Subramanian S, Hwang SJ, et al. Protein biomarkers of new-onset cardiovascular disease: prospective study from the systems approach to biomarker research in cardiovascular disease initiative. *Arterioscler Thromb Vasc Biol* 2014;34:939–945
32. Nordestgaard BG, Adourian AS, Freiberg JJ, Guo Y, Muntendam P, Falk E. Risk factors for near-term myocardial infarction in apparently healthy men and women. *Clin Chem* 2010;56:559–567
33. Contois JH, McConnell JP, Sethi AA, et al.; AACC Lipoproteins and Vascular Diseases Division Working Group on Best Practices. Apolipoprotein B and cardiovascular disease risk: position statement from the AACC Lipoproteins and Vascular Diseases Division Working Group on Best Practices. *Clin Chem* 2009;55:407–419
34. Sukhorukov V, Gudelj I, Pučić-Baković M, et al. Glycosylation of human plasma lipoproteins reveals a high level of diversity, which directly impacts their functional properties. *Biochim Biophys Acta Mol Cell Biol Lipids* 2019;1864:643–653
35. Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* 2018;50:1505–1513
36. Tanigaki K, Sacharidou A, Peng J, et al. Hyposialylated IgG activates endothelial IgG receptor Fc γ RIIB to promote obesity-induced insulin resistance. *J Clin Invest* 2018;128:309–322
37. Ohtsubo K, Chen MZ, Olefsky JM, Marth JD. Pathway to diabetes through attenuation of pancreatic beta cell glycosylation and glucose transport. *Nat Med* 2011;17:1067–1075
38. Ohtsubo K, Takamatsu S, Gao C, Korekane H, Kurosawa TM, Taniguchi N. N-Glycosylation modulates the membrane sub-domain distribution and activity of glucose transporter 2 in pancreatic beta cells. *Biochem Biophys Res Commun* 2013;434:346–351
39. Buse MG. Hexosamines, insulin resistance, and the complications of diabetes: current status. *Am J Physiol Endocrinol Metab* 2006;290:E1–E8
40. Lux A, Yu X, Scanlan CN, Nimmerjahn F. Impact of immune complex size and glycosylation on IgG binding to human Fc γ Rs. *J Immunol* 2013;190:4315–4323