Quantitative levels of serum N-glycans in type 1 diabetes and their association with kidney disease

Quantitative levels of serum *N*-glycans in type 1 diabetes and their association with kidney disease

Marco Colombo[a,§], Akram Asadi Shehni[b,§], Ioanna Thoma[c], Stuart J. McGurnaghan[c], Luke A.K. Blackbourn[c], Hayden Wilkinson[b], Andrew Collier[d], Alan W. Patrick[e], John R. Petrie[f], Paul M. McKeigue[g], Radka Saldova[b,h,‡], Helen M. Colhoun[c,i,‡,*], on behalf of the Scottish Diabetes Research Network (SDRN) Type 1 Bioresource Investigators

[a]*Independent consultant, Lecco, Italy*
[b]*NIBRT GlycoScience Group, National Institute for Bioprocessing Research and Training, Dublin, Ireland*
[c]*MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road South, Edinburgh EH4 2XU, UK*
[d]*School of Health and Life Sciences, Glasgow Caledonian University, Glasgow, UK*
[e]*Royal Infirmary of Edinburgh, NHS Lothian, Edinburgh, UK*
[f]*Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK*
[g]*Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK*
[h]*UCD School of Medicine, College of Health and Agricultural Science, University College Dublin, Dublin, Ireland*
[i]*Public Health, NHS Fife, Kirkcaldy, UK*
[§]*These authors contributed equally to this work.*
[‡]*These authors contributed equally to this work.*

* Corresponding Author
Helen M. Colhoun - helen.colhoun@igmm.ed.ac.uk
Institute of Genetics and Molecular Medicine
Western General Hospital
Crewe Road
Edinburgh, UK EH4 2XU
Tel : +44 (0) 131 651 8770

Running title: Levels of *N*-glycans and diabetic kidney disease

Keywords: diabetic kidney disease / glycaemic control / *N*-glycans / type 1 diabetes / UPLC

Supplementary Data Included: Tables SI – SVIII

Abstract

We investigated associations of quantitative levels of *N*-glycans with haemoglobin A_{1c} (HbA_{1c}), renal function and renal function decline in type 1 diabetes. We measured 46 total *N*-glycan peaks (GPs) on 1565 serum samples from the Scottish Diabetes Research Network Type 1 Bioresource Study (SDRNT1BIO) and a pool of healthy donors. Quantitation of absolute abundance of each GP used 2AB-labelled mannose-3 as standard. We studied cross-sectional associations of GPs and derived measures with HbA_{1c}, albumin/creatinine ratio (ACR) and estimated glomerular filtration rate (eGFR), and prospective associations with incident albuminuria and final eGFR.

All GPs were 1.4 to 3.2 times more abundant in SDRTN1BIO than in the healthy samples. Absolute levels of all GPs were slightly higher with higher HbA_{1c}, with strongest associations for triantennary trigalactosylated disialylated, triantennary trigalactosylated trisialylated structures with core or outer arm fucose, and tetraantennary tetragalactosylated trisialylated glycans. Most GPs showed increased abundance with worsening ACR. Lower eGFR was associated with higher absolute GP level, most significantly with biantennary digalactosylated disialylated glycans with and without bisect, triantennary trigalactosylated trisialylated glycans with and without outer arm fucose, and core fucosylated biantennary monogalactosylated monosialylated glycans. Although several GPs were inversely associated prospectively with final eGFR, cross-validated multivariable models did not improve prediction beyond clinical covariates.

Elevated HbA_{1c} is associated with an altered *N*-glycan profile in type 1 diabetes. Although we could not establish GPs to be prognostic of future renal function decline independently of HbA_{1c}, further studies to evaluate their impact in the pathogenesis of diabetic kidney disease are warranted.

Introduction

In diabetes, end organ damage is clearly associated prospectively with the degree of glycaemia (Perkins et al. 2019). This has been shown robustly in several large prospective studies in types 1 and 2 diabetes. More definitive evidence comes from the Diabetes Control and Complications Trial (DCCT), in which reduction of glycaemia led to a reduced incidence of end organ damage in the kidney, the eye and the cardiovascular system (de Boer 2014).

Glucose-induced tissue damage has been proposed to occur through a range of mechanisms including increased flux of glucose through the hexosamine biosynthetic pathway (HBP) (Rudman et al. 2019),(Ryczko et al. 2016). Since increased flux of glucose through the HBP in diabetes should lead to increased levels of uridine diphosphate-*N*-acetyl glucosamine (the donor molecule for the enzymatic process of *N*-linked glycans on secreted proteins) (Taparra et al. 2016), we previously hypothesized that elevated haemoglobin A_{1c} (HbA_{1c}) in people with diabetes would be associated with altered circulating *N*-glycans detectable on serum proteins (Bermingham et al. 2018). Specifically, we hypothesized that there would be an increase in the more branched (tri- and tetra-antennary, A3 and A4) glycan structures with higher HbA_{1c}. Such alterations in glycans modify protein function in pathways relevant to diabetic kidney disease (DKD), including epidermal growth factor (EGF) receptor and transforming growth factor-beta (TGF-beta) signaling (Lau et al. 2007). Consistent with this hypothesis, we showed that there were cross-sectional associations of *N*-glycan profile with HbA_{1c}, albumin creatinine ratio (ACR) and estimated glomerular filtration rate (eGFR). Elevated HbA_{1c}, albuminuria and greater reduction in eGFR in retrospective data were associated with relatively more branched, galactosylated and sialylated structures.

In the present follow-on study we have extended our investigation of this hypothesis to include absolute (and not just relative) abundance by means of a recently established approach to assay the absolute quantitation of *N*-glycans based on using fluorophore 2-aminobenzamide-labelled mannose-3 (Man3-2AB) as a standard (Colhoun et al. 2018). We hypothesized that a relationship of absolute levels of *N*-glycans with kidney outcomes would be more specific evidence of a causal relationship, since downstream effects of altered proteins would be expected to relate more closely to absolute than to relative amounts. The increase in the absolute amount of a particular glycan may arise from

one or more sources, such as (1) the general upregulation of certain glycosylated proteins that preferentially display that particular glycan structure, (2) increased glycosylation site occupancy of nonregulated proteins, or (3) alterations of the cellular glycosylation machinery in favor of the observed glycan structure (Moh et al. 2015). This may or may not be reflected in a relative quantitation approach.

Further, in this study we investigated 46 peaks, giving us a more detailed description of the *N*-glycan profile compared to our previous study, where some of these peaks were pooled and 39 were analyzed (Bermingham et al. 2018). Secondly, we increased the sample size and broadened it to include people with a wider range of starting eGFRs in order to examine relationships early in the course of DKD. Thirdly, we extended our analyses to test the *prospective* association of *N*-glycans with incident albuminuria and final eGFR in the Scottish Diabetes Research Network Type 1 Bioresource (SDRNT1BIO), a large type 1 diabetes cohort. In parallel, the concentration of *N*-glycans human serum in a pool of healthy donors was quantified in 46 peaks.

Results

Table I shows the distribution of clinical characteristics of those included in the study by eGFR stage at time of measurement of *N*-glycans. As expected, those with worse renal function (lower eGFR) had longer diabetes duration, worse blood pressures and higher ACR, were more likely to be on anti-hypertension treatment and specifically on angiotensin-converting enzyme inhibitors (ACEi) or angiotensin receptor blockers (ARB), and their eGFRs were more likely to decline faster, as determined by their prospective slopes.

Altogether, 46 glycan peaks were measured (Supplementary Table I and Fig. 1a), and from these we derived 18 common features calculated as described by Saldova et al. (Saldova et al. 2014) (Supplementary Table II). Supplementary Table III shows the summary characteristics of *N*-glycan peaks measured in the study for both the SDRNT1BIO and normal human healthy serum (NHS) samples. In terms of the absolute quantitation, all *N*-glycan peaks were between 1.4 and 3.2 times more abundant in SDRTN1BIO than in the NHS samples, with GP40, GP2, GP4, GP38, GP35 and

GP32 presenting a ratio of at least 2.5. By far the most abundant peaks in SDRNT1BIO were GP25, GP19, GP22, GP34, GP27 and GP24, all of which had mean absolute levels of at least 3 ng/μL, while only GP25 exceeded the same threshold in the NHS samples. When expressed as percentage areas, the distribution of *N*-glycan peaks was instead generally comparable between SDRNT1BIO and NHS, with GP40, GP38, GP22, GP35 and GP34 being the most statistically different. Altogether, 29% of the relative distribution of the glycans was accounted for by GP25 alone in both sample sets, with GP19 being the next most abundant.

N-glycans show strong cross-sectional associations with HbA_{1c}, ACR and eGFR

Table II summarises the univariate associations of *N*-glycan peaks with study day HbA_{1c} reporting the 10 strongest associations that reached the Bonferroni-adjusted significance threshold. In total there were 14 out of 46 *N*-glycan peaks that showed highly significant associations with HbA_{1c}, alongside 5 derived measures. The complete set of univariate associations of absolute levels and those with percentage areas, i.e. relative *N*-glycan amounts, is shown in Supplementary Table IV. The absolute level of *N*-glycans of all types tended to show higher abundance with higher HbA_{1c}. These associations were strongest for the complex branched, trigalactosylated (G3) and trisialylated (S3) structures, and those structures with outer arm fucose, such that both their absolute and relative levels were higher with higher HbA_{1c} (Fig. 1b). Associations for relative amounts were similar to our previous study (Bermingham et al. 2018), in that the complex branched tri- and tetra-antennary structures were relatively more abundant when HbA_{1c} was higher and were relatively less abundant when HbA_{1c} was lower.

The second section in Table II shows the top 10 associations of *N*-glycans with ACR at study day (the complete set of univariate associations is shown in Supplementary Table V). Most *N*-glycans were more abundant with worse ACR, the differences being highly significant for 23 *N*-glycans. The increase was in both the relative distribution of several *N*-glycans as well as their absolute levels. The complex branched *N*-glycans with more galactosylation and sialylation were more abundant when ACR was worse (Fig. 1b). The same pattern that we saw in our previous study of relative amounts was replicated. The *N*-glycans associated with ACR were mostly the same as those associated with HbA_{1c}.

The third section of Table II shows the associations between *N*-glycans with study day eGFR (the complete set of univariate associations is shown in Supplementary Table VI). It can be seen that, in general, the lower the eGFR, the higher the *N*-glycan level for all *N*-glycans examined in terms of absolute amounts while the complex branched glycans increased to a greater extent than did the other glycans (Fig. 1b). However, the pattern of the same *N*-glycans associated with worse glycaemia being increased was not quite as clear as it was for ACR. While higher branched complex *N*-glycans GP32, GP37 and those with outer arm fucose were associated with eGFR as with albuminuria and HbA_{1c}, so were some of the biantennary disialylated *N*-glycans.

Some differences were observed in glycan associations among HbA_{1c}, ACR and eGFR (Fig. 1b). Most importantly, some glycan peaks which were positively associated with ACR and/or negatively associated with eGFR were not associated with HbA_{1c}. Those included biantennary disialylated glycans (GP25 and GP26) that were associated with both ACR and eGFR. Also, oligomannose, core fucosylated mono- and bi-antennary glycans containing up to two sialic acids (GP5, GP6, GP8-GP11, GP14, GP17, GP27) as well as tetraantennary tetrasialylated glycans (GP42-GP45) were only associated with eGFR. In addition, monoantennary monosialylated and biantennary non sialylated glycan-containing peak (GP12) was only associated with ACR (Fig. 1b).

Supplementary Table VII presents a comparison of the significant associations of relative concentrations of *N*-glycan peaks with HbA_{1c} and ACR obtained in this study to those previously reported in Bermingham et al. (Bermingham et al. 2018). Given the remarkable agreement between the two studies for those peaks whose structures can be matched unequivocally, our results support previous associations with type 1 diabetes markers.

N-glycans are prospectively associated with incident albuminuria and eGFR progression

We analysed 1533 individual samples for models of final eGFR and 1475 samples for progression to final eGFR < 45 mL/min/1.73 m^2 (78 events). There were 283 incident cases of albuminuria in our analysis in the set of 1278 individuals with no albuminuria at study day.

Table III shows the top 10 univariate associations of *N*-glycans with incident albuminuria, final eGFR and progression to eGFR < 45 mL/min/1.73 m^2, both with and without adjustment for HbA_{1c} as

well as age, sex, diabetes duration and length of follow-up (and study day eGFR for the two prospective eGFR outcomes).

For incident albuminuria, odds ratios strongly above 1 were seen for both the absolute and relative abundance of larger, more branched *N*-glycans. Significant associations were found for GP32, GP40, GP31 and GP39; only the association with GP32 was independent of further adjustment for HbA_{1c} (Table III, first section). GP32 mostly contains A3G3S3, but this structure is common to other peaks that were not significantly associated with incident albuminuria. However, as reported in Saldova et al. (Saldova et al. 2014), GP32 contains in lower abundance a unique glycan structure of A3F1G3S2 (triantennary outer arm fucosylated trigalactosylated disialylated glycan) as well as even lower amounts of A4G4S1 (tetraantennary tetragalactosylated monosialylated glycan). A3F1G3S2 might be a potential marker for incident albuminuria, as A4G4S1 is also found in GP30, which was not significant for incident albuminuria.

There was a clear pattern of higher absolute levels of all *N*-glycans being associated with lower final eGFR in models adjusted for baseline eGFR and other covariates (Table III, second section), with several associations reaching Bonferroni-corrected significance. The absolute level of complex branched *N*-glycans (GP37, GP31, GP34, GP29, GP32) and the summary measures corresponding to more branched (A3, A3A4), trigalactosylated (G3) and trisialylated (S3) structures and those with outer arm fucose were significantly associated with lower final eGFR adjusted for baseline eGFR, age, sex and diabetes duration. On adjustment for HbA_{1c}, most of these associations remained significant. When adjusted for further covariates (body mass index, diastolic and systolic blood pressure, HDL-cholesterol, total cholesterol, smoking status, ACR, and being on ACEi or ARB), the regression coefficients were much smaller, and no association reached statistical significance (data not shown). Interestingly, GP18 and GP20 were negatively associated with final eGFR but not with study day eGFR (Fig. 1b). GP18 contains core fucosylated bisecting biantennary monogalactosylated monosialylated glycan, while GP20 contains bisecting biantennary digalactosylated monosialylated glycan.

For the binary progression outcome to final eGFR < 45 mL/min/1.73 m^2, there was a pattern broadly similar to that for final eGFR. Bonferroni-adjusted significant associations were seen for 7 of

the *N*-glycans and derived measures tested (Table III, third section). For this less prevalent outcome, adjustment for HbA_{1c} reduced the coefficients so that they no longer reached significance. However, it is worth mentioning that in GP46, the only peak containing three outer arm fucoses, there was only a minimal reduction in its effect size upon adjustment for HbA_{1c}. This might suggest the importance of highly outer arm fucosylated glycan for this outcome (final eGFR < 45 mL/min/1.73 m^2).

Performance of N-glycans as predictive biomarkers for albuminuria and eGFR decline

Supplementary Table VIII summarises the predictive performance of adding either the set of *N*-glycan peaks, or the derived measures, or both sets simultaneously, to baseline models containing age, sex, diabetes duration, study day eGFR, HbA_{1c} and length of follow-up for prediction of last available eGFR, progression to eGFR < 45 mL/min/1.73 m^2, and incident albuminuria. Whether considered in terms of r^2, area under the receiver operating characteristic curve (AUC) or expected information for discrimination Λ, despite showing some strong univariate associations, the *N*-glycans do not contain any additional predictive information beyond these other clinical covariates, as the cross-validated performance for all outcomes examined is effectively unchanged when they are added to the models.

Discussion

To our knowledge, this is the most comprehensive study of absolute concentrations of human serum *N*-glycans. Previous study of Kita et al. quantified 34 major *N*-glycans from human serum (Kita et al. 2007). Here we measured and presented in detail quantification of 46 *N*-glycan peaks by HILIC-UPLC containing 165 glycans that were previously presented by Saldova et al. (Saldova et al. 2014). In addition, we have applied this technology to investigate associations of these serum *N*-glycans with glycaemic control and renal complications in type 1 diabetes. We employed a recently established approach to derive the absolute level of *N*-glycans as the assay methodology previously adopted in Bermingham et al. (Bermingham et al. 2018) could only report relative quantitation. Here we showed that elevated HbA_{1c} is strongly associated with complex branched *N*-glycans circulating in absolute and not just relative terms. Furthermore, we showed in prospective analyses that lower eGFR at

follow-up (adjusted for study day eGFR and the other covariates) and incident albuminuria were associated with a greater abundance of complex branched structures as well as more abundant branched structures with outer arm fucosylation. Although further adjustment for HbA_{1c} reduced the associations slightly, they remained significant for final eGFR.

The role of *N*-linked glycosylation in diabetes has been recently summarized in the review by Rudman et al. (Rudman et al. 2019). The striking associations of *N*-glycan patterns with glycaemia suggest that the role of altered protein modification with *N*-glycans in downstream tissue damage should be further explored. We first investigated this in our previous work, where we considered how relative concentrations of 39 *N*-glycan peaks associated cross-sectionally to HbA_{1c}, ACR and eGFR slopes (Bermingham et al. 2018). A limitation of that approach is that *N*-glycans are not directly quantified, and thus it is impossible to discern whether an increase in the relative concentration of a glycan peak is due to an increment of the absolute level of that peak rather than to a reduction of other peaks. Here we demonstrated for the first time that the absolute levels of complex branched *N*-glycans is increased in the presence of worse glycaemia in type 1 diabetes. We had hypothesized that this would be the case because metabolic flux through the HBP increases *N*-glycan branching (Lau et al. 2007). Our demonstration here of a prospective association of *N*-glycan patterns with renal outcomes independently of baseline HbA_{1c} and eGFR is consistent with a causal association and provides stronger evidence than our previous cross-sectional report (Bermingham et al. 2018). A possible mechanism is disruption of the glycocalyx by altered *N*-glycan branching and levels of sialylation and galactosylation. It is well established that abnormalities of the glycocalyx occur early in diabetic nephropathy (Fu et al. 2015; Yilmaz et al. 2019) and there is extensive literature suggesting that *N*-glycans are of pivotal importance to the maintenance of the proper functioning of the glycocalyx. Branched *N*-glycans found on transmembrane proteins are the major ligands for galectins (Patnaik et al. 2005). The affinity for galectins varies with the degree of branching; tri- and tetra-antennary *N*-glycans bind to the galectins and form a molecular lattice at the cell surface. This regulates surface levels of important glycoproteins in the glycocalyx (Boscher et al. 2011). There is also considerable evidence that *N*-glycan branching differentially regulates epidermal growth factor receptor and TGF-beta signalling (Partridge et al. 2004), both of which pathways are known to be

highly relevant and disruptive in DKD (Böttinger 2007; Harskamp et al. 2016). The epidemiological data provided by our study add support to a role for more branched *N*-glycan structures in mediating glycaemia related renal tissue damage. However, an effect of modifying *N*-glycan levels on relevant pathologic features in *in vitro* and *in vivo* models is of course needed to provide more convincing evidence of causality.

Inflammation is important for diabetic nephropathy progression, and anti-inflammatory drugs could protect against nephropathy (Moreno et al. 2018; Sabapathy et al. 2019). Progressive renal decline (eGFR) is associated with the presence of specific pro-inflammatory cytokines (Bjornstad et al. 2014). Glycosylation on serum proteins comes mostly from IgG from plasma cells and liver acute phase proteins (Hamfjord et al. 2015). Pro-inflammatory cytokines were shown to induce increase in 2, 6 linked sialic acid in human primary cells and sialyl Lewis X epitopes (sLex, containing outer arm fucose), branching and sialylation in human chronic inflammation (Hanasaki et al. 1994; Arnold et al. 2008). These cytokines also cause increase in branching and sialylation on serum liver glycoproteins in mouse model (Saldova et al. 2013). Anti-TNF therapy, that targets a pro-inflammatory cytokine, in rheumatoid arthritis led to decreased sialylation on serum glycoproteins and increased galactosylation and sialylation on IgG (IgG mostly contains biantennary *N*-glycans) (Stadlmann et al. 2008), (Collins et al. 2013; Shade and Anthony 2013). It has been demonstrated that increased branching, galactosylation and sialylation on certain acute phase proteins may lead to a longer half-life of these proteins (Saldova et al. 2008). Association of increased branching, sialylation and sLex glycans with HbA_{1c}, ACR, and eGFR decline in this study might be results of the inflammatory processes involved in diabetic nephropathy progression.

Our primary interest was to establish if absolute levels of *N*-glycans are prospectively associated with renal damage in diabetes, supporting a causal role. Moreover, we wanted to quantify if *N*-glycans would be of practical use as predictive biomarkers as *N*-glycan patterns are being developed as predictive biomarkers in other diseases (Reily et al. 2019). Despite the significant prospective associations we demonstrated, modelling shows that *N*-glycans do not usefully improve prediction of renal outcomes when considered above clinical covariates. However, associations of absolute levels of *N*-glycans with prospective outcomes is consistent with a causal explanation of their role in renal

function decline in individuals with type 1 diabetes, and provides more convincing evidence than could be obtained in our previous work (Bermingham et al. 2018).

The strengths of this study include the large sample size, the use of a recently established absolute quantitative method for *N*-glycans, the extensive covariate adjustment, the appropriate prediction modelling using cross-validation and novel methods to avoid overfitting. One of the limitations of this study is the relatively short duration of follow-up at present. For instance, only in 78 participants a loss of renal function corresponding to eGFR falling below 45 mL/min/1.73 m^2 had occurred during follow-up, thereby reducing the power to demonstrate improvements in prediction of renal outcomes.

While knowing the exact glycan quantities can provide additional information such as the total amount of glycosylation on the carrier proteins and how it varies in biology, this additional information could potentially hold significant biological importance if it were complemented with proteome or glycoproteome data (Moh et al. 2015). This limitation could be addressed in future studies considering how these changes in the glycan signatures relate to changes in the individual glycoproteins such as acute phase proteins or immunoglobulins in diabetic patients. Currently, similar conclusions were reached using either relative (in line with previously published Bermingham et al. (Bermingham et al. 2018)) or absolute quantitation.

Our findings show that altered *N*-glycan profile on circulating proteins is a feature of worse glycaemic control in type 1 diabetes, and the *N*-glycan profile predicts incident albuminuria and worse eGFR progression. Further studies to evaluate the potential impact of altered *N*-glycans in the pathogenesis of DKD are warranted.

Materials and Methods

Study population

The Scottish Diabetes Research Network Type 1 Bioresource Study (SDRNT1BIO) is a population- representative cohort study of people with type 1 diabetes aged 16 years and older who were enrolled between 1 December 2010 and 29 November 2013. The study design and data collection procedures have been described in detail (Akbar et al. 2017). Participants are

representative of the total adult population with type 1 diabetes in Scotland with regard to clinical history as well as clinical and demographic characteristics. In brief, 6127 participants attended a single visit at which questionnaires were completed, physical measures taken and non-fasting blood and random urine samples were acquired. The study data were linked retrospectively and prospectively to other routine data sources, including electronic health records for diabetes and laboratory data.

The study was approved by the Tayside Research Ethics Committee (reference 10/S1402/43) and conducted according to the principles of the Declaration of Helsinki. Written consent was obtained from the participants.

Acquisitions of HbA_{1c}, ACR and eGFR data

For all SDRNT1BIO participants, we linked HbA_{1c}, ACR and eGFR and serum creatinine measurements from the Scottish Care Information-Diabetes Collaboration database (SCI-Diabetes) to participant study day data. This database captures all of the healthcare systems laboratory measures on these analytes in people with type 1 diabetes. Data both retrospective and prospective to study day were linked up to end 2018. We also directly measured these analytes in the study day samples. Baseline HbA_{1c} was defined as the measurement closest to, and up to 18 months before, the date of recruitment.

eGFR was calculated from serum creatinine data with the CKD-EPI equation (Levey et al. 2009) using serum creatinine values directly measured and retrieved retrospectively and prospectively from medical records (with a median of 22 measurements per person, of which 11 were collected during a median of 5.4 years preceding the enrolment day, and 11 from then onwards). To validate the use of the eGFR from the electronic record, we confirmed that the health record–derived eGFR at time of sampling was highly correlated ($r > 0.97$) with eGFR computed from directly-measured serum creatinine. Initiation of renal replacement therapy (RRT) was considered to indicate an achieved eGFR of 10 mL/min/1.73 m^2 and all subsequent readings were censored. Study day eGFR was computed from the serum creatinine sample obtained when the participant was enrolled into the study. Last available eGFR (final eGFR) was defined as the median eGFR reading of the last six months of follow-up (for 27 participants with follow-up shorter than six months, the computation of this median eGFR reading included the study day measurement). We then defined a binary outcome of

progression to a final eGFR < 45 mL/min/1.73 m^2. We used the KDIGO 2012 definitions of chronic kidney disease (CKD) stages (KDIGO Work Group 2013) according to ranges of eGFR in mL/min/1.73 m^2: G1: > 90; G2: 60–90; G3: 30–60; G4: 15–30.

ACR was measured in paired urine samples, with the first taken at study day and the second several days later using the ADVIA 2400 immunoturbidimetric method for albumin and the ADVIA 2400 enzymatic method for creatinine (Siemens, Munich, Germany). These data were used for adjusting for ACR in the analyses here. In addition, longitudinal urinary ACR data was captured from the routine clinical laboratory data. Clinical record data close to study day were highly correlated (r = 0.73) with the direct measurement. At any time point, albuminuric status was defined based on the most recent available albuminuria measurement provided there was no contradictory record of that stage in the preceding or subsequent 90 days, i.e. someone who transited from normo- to microalbuminuria but then had another normoalbuminuria measurement within 90 days was assigned as having been normoalbuminuric across that period, such that transient changes in albuminuria readings were ignored. Participants were assigned a label of normo-, micro- or macroalbuminuric according to their ACR falling in the intervals 0–3.39, 3.39–33.9, or > 33.9 mg/mmol, based on 2 out of 3 consecutive measurements before baseline.

Retrospective and prospective covariate data including drug exposures were obtained from the electronic healthcare record SCI-Diabetes as described previously.

Acquisition of N-glycans data

From the total SDRNT1BIO sample bank we sampled a proportion of participants (n=1565) with varying sampling fractions depending on CKD stage as follows: 20% of those at stage G1 (n=827), 30% of those at stage G2 (n=557), 60% of those at stage G3 or worse (n=181). For 16 participants, a blinded duplicate sample was also measured. Quality control samples run with each set of diabetic patient samples were water (negative control) and a normal human healthy serum (NHS), a pool of serum from 100 healthy donors (Saldova et al. 2014).

N-glycans were released from SDRNT1BIO participant serum samples and NHS controls (5 µL) using the high-throughput automated method previously described by Stöckmann et al. (Stöckmann et al. 2015). Briefly, the sample denaturation was carried out with dithiothreitol in ammonium bicarbonate

at 65 °C, alkylation with iodoacetamide, and trypsin solution was added (10 µL at 40,000 U/mL) at 40 °C. Finally, *N*-glycans were released from the protein backbone enzymatically via PNGase F (Prozyme peptide *N*-glycanase F, GKE-5006D, 10 µL per well, 0.5 mU in 1 M ammonium bicarbonate, pH 8.0). The glycans from half of the enzymatically treated samples were immobilized on solid supports hydrazide beads (ThermoScientific, Waltham, MA, USA) in 96-well chemically inert filter plates (Millipore Solvinert, hydrophobic polytetrafluoroethylene membrane, 0.45 µm pore size, Billerica, MA, USA), and excess reagents were removed on the robotic vacuum manifold. Glycans were then released from the solid supports and labelled with the 2AB. Next, glycans were cleaned up using 96-well chemically inert filter plates and HyperSepDiol Solid Phase Extraction cartridges (ThermoScientific, Waltham, MA, USA) (Stöckmann et al. 2013; Stöckmann et al. 2015).

Quantitation was done by spiking all samples with Man3-2AB (Carbosynth Ltd, Berkshire, UK) standard according to Colhoun et al. (Colhoun et al. 2018) at the concentration of 1.67 ng.

Ultra performance liquid chromatography (UPLC)

2AB derivatized *N*-glycans were separated by UPLC with fluorescence detection on a Waters Acquity UPLC instrument consisting of a binary solvent manager, sample manager and fluorescence detector under the control of Empower 3 chromatography workstation software (Waters, Milford, MA, USA). The HILIC separations were performed using a Waters Ethylene Bridged Hybrid (BEH) Glycan column, 150 x 2.1 mm, 1.7 µm BEH particles, with 50 mM ammonium formate, pH 4.4, as solvent A and MeCN as solvent B. The 30 minute method was used with a linear gradient of 30-47% with buffer A at 0.56 mL/min flow rate for 23 minutes followed by 47-70% A for 1 min and finally reverting back to 30% A in another 1 min and to complete the run (Saldova et al. 2014). An injection volume of 10 µL sample prepared in 70% v/v MeCN was used throughout. Samples were maintained at 5 °C prior to injection, while separation was carried out at 40 °C. The fluorescence detection excitation/emission wavelengths were ex = 330 nm and em = 420 nm, respectively. The system was calibrated using an external standard of hydrolyzed and 2AB-labeled glucose oligomers to create a dextran ladder, as described previously (Royle et al. 2006).

Total serum *N*-glycome from each patient was separated into 46 peaks (containing group of

glycans, according to Saldova et al. (Saldova et al. 2014)), of individual relative proportion to total 100% peak area (relative quantitation of glycans, proportion of certain glycans to total glycome) or plotted as absolute levels calculated as proportion of the glycans to the Man3-2AB standard. Glycan peaks were expressed as absolute levels in ng/μL in volume of serum as well as percentage areas. Percentage areas were computed as the ratio between each peak area and the sum of all peak areas for each participant multiplied by 100; they express the relative abundance of each *N*-glycan peak with respect to all other peaks. Measurements from participants with blinded duplicate samples were averaged. The *N*-glycan peaks were complemented by 18 derived summary measures that aggregated the sums of peaks having features related to antennary structure, galactosylation, fucosylation, or sialylation, as presented in (Saldova et al. 2014).

Statistical analysis

For models of final eGFR, we removed 32 observations with less than 1 year of follow-up data, and thus the dataset contained 1533 individuals. In models of progression to final eGFR < 45 mL/min/1.73 m^2, we further removed 58 observations that already had an eGFR below 45 mL/min/1.73 m^2 at study day. Incident albuminuria was defined as being normoalbuminuric at baseline and having two urine samples assay results consistent with micro- or macroalbuminuria at any time during follow-up, or one if that was the last. Models for incident albuminuria contained 1278 individuals with no prevalent albuminuria at study day. All continuous variables were Gaussianized and standardized to zero mean and unit standard deviation.

Univariate analyses

Analyses were performed using *N*-glycan levels in absolute units as well as percentage areas. Glycans were evaluated independently in linear and logistic regression models as appropriate. For associations with HbA_{1c}, we considered each glycan peak or derived measure as the outcome variable, and tables report the regression coefficient for HbA_{1c} after adjusting for age, sex and diabetes duration. For all other models of associations, each glycan peak or derived measure was used as a predictor, and models were adjusted for age, sex, diabetes duration; prospective models for incident albuminuria were further adjusted for length of follow-up, and models for final eGFR and

incident final eGFR < 45 mL/min/1.73 m^2 were further adjusted by study day eGFR and length of follow-up. In these analyses we examined associations both with and without HbA_{1c} in the model. Associations were declared significant at a Bonferroni-corrected $p = 0.05/128 = 3.9 \times 10^{-4}$.

Multivariate analysis

To examine whether panels of *N*-glycans are useful as biomarkers to improve prediction of incident albuminuria or eGFR progression we used a penalized Bayesian approach based on hierarchical shrinkage (Colombo et al. 2019). In this analysis, *N*-glycans were assigned a regularized horseshoe prior (Piironen and Vehtari 2017), while for covariates we employed a weakly informative Gaussian prior. This approach induces a strong penalization of biomarker effects, which is particularly beneficial when examining a large number of potential predictors, as it helps in reducing overfitting when learning regression coefficients. Therefore, the coefficients of the most informative *N*-glycans can be large while the remaining coefficients can be shrunk toward zero. The models were fitted using the R package hsstan (version 0.8: https://CRAN.R-project.org/package=hsstan).

We evaluated the predictive performance of the models including *N*-glycans on withdrawn data using 10-fold cross-validation, and compared models including *N*-glycans to baseline models that only contain clinical covariates. We reported the difference in log-likelihood (computed on the test observations from 10-fold cross-validation, and expressed in natural log units) between the model with *N*-glycans and the model with only the clinical covariates. It can be shown that a difference in test log-likelihood of at least 2.7 natural log units between nested models would be required to report a significance level of $p < 0.01$ (Stone 1977). For linear regression models we computed the r^2 as the squared Pearson correlation coefficient between observed and predicted outcome. For logistic regression models we reported the AUC and the expected information for discrimination Λ expressed in bits (McKeigue 2019). By comparing the Λ obtained from a model that contains clinical covariates and potential biomarkers to the Λ from a model containing only the set of clinical covariates we can quantify the incremental contribution of potential biomarkers to the predictive performance of the model, as this difference captures the amount of additional information that biomarkers contain over and beyond the initial set of clinical covariates.

Funding

This study was supported by funding from Juvenile Diabetes Research Foundation (Ref. 3-SRA-2016- 332-M-R) Chief Scientist Office (Ref. ETM/47); Diabetes UK (Ref. 10/0004010); in-kind contribution from Scottish Diabetes Research Network. RS acknowledges funding from the Science foundation Ireland Starting Investigator Research grant (SFI SIRG) under grant number 13/SIRG/2164.

Acknowledgements

We thank the SDRN Type 1 Bioresource investigators: J. Chalmers (Diabetes Centre, Victoria Hospital, UK), A. Collier (School of Health and Life Sciences, Glasgow Caledonian University, Glasgow, UK), C. Fischbacher (Information Services Division, NHS National Services Scotland, Edinburgh, UK), F. Green (Research & Development Support Unit, Dumfries & Galloway Royal Infirmary, Dumfries, UK), R. Lindsay (British Heart Foundation Glasgow Cardiovascular Research Centre, University of Glasgow, UK), J. McKnight (Western General Hospital, NHS, UK), S. MacRury (Highland Diabetes Institute, Raigmore Hospital, NHS Highland, Inverness, UK), C. Palmer (Cardiovascular and Diabetes Medicine, University of Dundee, UK), A. Patrick (Royal Infirmary of Edinburgh, NHS Lothian, UK), D. Pearson (JJR Macleod Centre for Diabetes, Endocrinology and Metabolism, Aberdeen Royal Infirmary, UK), J. Petrie (Institute of Cardiovascular & Medical Sciences, University of Glasgow, UK) and S. Thekkepat (David Matthews Diabetes Centre, Monklands Hospital, UK). We thank the nurses across the participating sites in SDRN Type 1 Bioresource, and those with diabetes who took part.

SDRNT1BIO was originally set up under Ethics ref 10/S1402/43, PAC 15/13, and is now running under PBPP ref. 1819-0315.

Data availability

The *N*-glycan measurements for the normal human serum samples and the SDRNT1BIO participants that entered the study are available for download from https://doi.org/10.7488/ds/2856. We do not have governance permissions to share individual level data on which these analyses were

conducted since they derive from clinical record data. However, for any *bona fide* requests to audit the validity of the analyses, the verifiable research pipeline which we operate means that one can request to view the analyses being run and the same tabulations resulting. We are also happy to share summary statistics for those wishing to conduct meta-analyses with other studies.

Conflict of interest statement

HMC receives research support and honorarium and is also a member of the advisory panels and speaker's bureaus for Sanofi Aventis, Regeneron, and Eli Lilly. HMC has been a member of DSMB of the Advisory Panel for the CANTOS Trial (Novartis Pharmaceuticals). HMC also receives or has recently received a non-binding research support from Pfizer Inc., and AstraZeneca LP. HMC is a shareholder of Roche Pharmaceuticals and Bayer.

MC, AAS, IT, SJM, LAKB, HW, AC, AWP, JRP, PMM and RS declare that there is no duality of interest associated with their contribution to this manuscript.

Abbreviations

ACEi, angiotensin-converting enzyme inhibitors; ACR, albumin/creatinine ratio; ARB, angiotensin receptor blockers; AUC, area under the receiver operating characteristic curve; CKD, chronic kidney disease; CKD-EPI, Chronic Kidney Disease Epidemiology Collaboration; DKD, diabetic kidney disease; eGFR, estimated glomerular filtration rate; HbA_{1c}, haemoglobin A_{1c}; HBP, hexosamine biosynthetic pathway; HILIC, hydrophilic interaction liquid chromatography; NHS, normal human healthy serum; RRT, renal replacement therapy; SDRNT1BIO, Scottish Diabetes Research Network Type 1 Bioresource; UPLC, ultra performance liquid chromatography.

References

Akbar T, McGurnaghan S, Palmer CNA, Livingstone SJ, Petrie J, Chalmers J, Lindsay RS, McKnight JA, Pearson DWM, Patrick AW, et al. 2017. Cohort Profile: Scottish Diabetes Research Network Type 1 Bioresource Study (SDRNT1BIO). *Int J Epidemiol*. 46(3):796–796i. doi:10.1093/ije/dyw152.

Arnold JN, Saldova R, Hamid UMA, Rudd PM. 2008. Evaluation of the serum N-linked glycome for the diagnosis of cancer and chronic inflammation. *Proteomics*. 8(16):3284–3293.

Bermingham ML, Colombo M, McGurnaghan SJ, Blackbourn LAK, Vučković F, Pučić Baković M, Trbojević-Akmačić I, Lauc G, Agakov F, Agakova AS, et al. 2018. N-Glycan Profile and Kidney Disease in Type 1 Diabetes. *Diabetes Care*. 41(1):79–87.

Bjornstad P, Cherney D, Maahs D. 2014. Early diabetic nephropathy in type 1 diabetes: new insights. *Current Opinion in Endocrinology & Diabetes and Obesity*. 21(4):279–286.

de Boer IH for the DRG. 2014. Kidney Disease and Related Findings in the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Study. *Diabetes Care*. 37(1):24–30.

Boscher C, Dennis JW, Nabi IR. 2011. Glycosylation, galectins and cellular signaling. *Current Opinion in Cell Biology*. 23(4):383–392.

Böttinger EP. 2007. TGF-β in Renal Injury and Disease. *Seminars in Nephrology*. 27(3):309–320.

Colhoun HO, Treacy EP, MacMahon M, Rudd PM, Fitzgibbon M, O'Flaherty R, Stepien KM. 2018. Validation of an automated ultraperformance liquid chromatography IgG N-glycan analytical method applicable to classical galactosaemia. *Annals of Clinical Biochemistry*. 55(5):593–603.

Collins ES, Galligan MC, Saldova R, Adamczyk B, Abrahams JL, Campbell MP, Ng C-T, Veale DJ, Murphy TB, Rudd PM, et al. 2013. Glycosylation status of serum in inflammatory arthritis in response to anti-TNF treatment. *Rheumatology*. 52(9):1572–1582.

Colombo M, Valo E, McGurnaghan SJ, Sandholm N, Blackbourn LAK, Dalton RN, Dunger D, Groop P-H, McKeigue PM, Forsblom C, et al. 2019. Biomarker panels associated with progression of renal disease in type 1 diabetes. *Diabetologia*. 62(9):1616–1627. doi:10.1007/s00125-019-4915-0. [accessed 2019 Sep 25]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6677704/.

Fu J, Lee K, Chuang PY, Liu Z, He JC. 2015. Glomerular endothelial cell injury and cross talk in diabetic kidney disease. *American Journal of Physiology-Renal Physiology*. 308(4):F287–F297.

Hamfjord J, Saldova R, Stöckmann H, Sandhu V, Bowitz Lothe IM, Buanes T, Lingjærde OC, Labori KJ, Rudd PM, Kure EH. 2015. Serum N-glycome characterization in patients with resectable periampullary adenocarcinoma. *Journal of Proteome Research*. 14(12):5144–5156.

Hanasaki K, Ajit V, Ivan S, Bevilacqua MP. 1994. Cytokine-induced beta-galactoside alpha-2,6-sialyltransferase in human endothelial cells mediates alpha 2,6-sialylation of adhesion molecules and CD22 ligands. *Journal of Biological Chemistry*. 269(14):10637–10643.

Harskamp LR, Gansevoort RT, van Goor H, Meijer E. 2016. The epidermal growth factor receptor pathway in chronic kidney diseases. *Nature Reviews Nephrology*. 12(8):496–506.

KDIGO Work Group. 2013. KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. *Kidney Intenational Supplements*. 3:1–150.

Kita Y, Miura Y, Furukawa J, Nakano M, Shinohara Y, Ohno M, Takimoto A, Nishimura S-I. 2007. Quantitative Glycomics of Human Whole Serum Glycoproteins Based on the Standardized Protocol for Liberating N-Glycans. *Molecular & Cellular Proteomics*. 6(8):1437–1445. doi:10.1074/mcp.T600063-MCP200. [accessed 2020 May 27]. https://www.mcponline.org/content/6/8/1437.

Lau KS, Partridge EA, Grigorian A, Silvescu CI, Reinhold VN, Demetriou M, Dennis JW. 2007. Complex N-Glycan Number and Degree of Branching Cooperate to Regulate Cell Proliferation and Differentiation. *Cell*. 129(1):123–134.

Levey AS, Stevens LA, Schmid CH, Zhang Y, Castro I Alejandro F, Feldman HI, Kusek JW, Eggers P, Van

Lente F, Greene T, et al. 2009. A new equation to estimate glomerular filtration rate. *Annals of Internal Medicine*. 150(9):604–612.

McKeigue P. 2019. Quantifying performance of a diagnostic test as the expected information for discrimination: Relation to the C-statistic. *Statistical methods in medical research*. 28(6):1841–1851. doi:10.1177/0962280218776989.

Moh ESX, Thaysen-Andersen M, Packer NH. 2015. Relative versus absolute quantitation in disease glycomics. *PROTEOMICS – Clinical Applications*. 9(3–4):368–382. doi:10.1002/prca.201400184. [accessed 2020 Oct 29]. https://onlinelibrary.wiley.com/doi/abs/10.1002/prca.201400184.

Moreno JA, Gomez-Guerrero C, Mas S, Belen Sanz A, Lorenzo O, Ruiz-Ortega M, Opazo L, Mezzano S, Egido J. 2018. Targeting inflammation in diabetic nephropathy: a tale of hope. *Expert Opinion on Investigational Drugs*. 27(11):917–930.

Partridge EA, Le Roy C, Di Guglielmo GM, Pawling J, Cheung P, Granovsky M, Nabi IR, Wrana JL, Dennis JW. 2004. Regulation of Cytokine Receptors by Golgi N-Glycan Processing and Endocytosis. *Science*. 306(5693):120–124.

Patnaik SK, Potvin B, Carlsson S, Sturm D, Leffler H, Stanley P. 2005. Complex N-glycans are the major ligands for galectin-1, -3, and -8 on Chinese hamster ovary cells. *Glycobiology*. 16(4):305–317.

Perkins BA, Bebu I, de Boer IH, Molitch M, Tamborlane W, Lorenzi G, Herman W, White NH, Pop-Busui R, Paterson AD, et al. 2019. Risk Factors for Kidney Disease in Type 1 Diabetes. *Diabetes Care*. 42(5):883–890.

Piironen J, Vehtari A. 2017. Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*. 11:5018–5051.

Reily C, Stewart TJ, Renfrow MB, Novak J. 2019. Glycosylation in health and disease. *Nature Reviews Nephrology*. 15(6):346–366.

Royle L, Radcliffe CM, A DR, Rudd PM. 2006. Detailed structural analysis of N-glycans released from glycoproteins in SDS-PAGE gel bands using HPLC combined with exoglycosidase array digestions. In: I B, editor. Glycobiology protocols. *Methods in molecular biology*, vol 347. Humana Press.

Rudman N, Gornik O, Lauc G. 2019. Altered N-glycosylation profiles as potential biomarkers and drug targets in diabetes. *FEBS Letters*. 593(13):1598–1615. doi:10.1002/1873-3468.13495. [accessed 2020 May 27]. https://febs.onlinelibrary.wiley.com/doi/abs/10.1002/1873-3468.13495.

Ryczko MC, Pawling J, Chen R, Abdel Rahman AM, Yau K, Copeland JK, Zhang C, Surendra A, Guttman DS, Figeys D, et al. 2016. Metabolic Reprogramming by Hexosamine Biosynthetic and Golgi N-Glycan Branching Pathways. *Scientific Reports*. 6(1):23043. doi:10.1038/srep23043. [accessed 2020 Oct 29]. https://www.nature.com/articles/srep23043.

Sabapathy V, Stremska ME, Mohammad S, Corey RL, Sharma PR, Sharma R. 2019. Novel immunomodulatory cytokine regulates inflammation, diabetes, and obesity to protect from diabetic nephropathy. *Frontiers in Pharmacology*. 10(572).

Saldova R, Asadi Shehni A, Haakensen VD, Steinfeld I, Hilliard M, Kifer I, Helland Å, Yakhini Z, Børresen-Dale A-L, Rudd PM. 2014. Association of N-glycosylation with breast carcinoma and systemic features using high-resolution quantitative UPLC. *Journal of Proteome Research*. 13(5):2314–2327.

Saldova R, Piccard H, Pérez-Garay M, Harvey DJ, Struwe WB, Galligan MC, Berghmans N, Madden SF, Peracaula R, Opdenakker G, et al. 2013. Increase in sialylation and branching in the mouse serum N-glycome correlates with inflammation and ovarian tumour progression. *PLOS ONE*. 8(8):1–11.

Saldova R, Wormald MR, Dwek RA, Rudd PMR. 2008. Glycosylation changes on serum glycoproteins in ovarian cancer may contribute to disease pathogenesis. *Disease Markers*. 25(4–5):219–232.

Shade K-TC, Anthony RM. 2013. Antibody glycosylation and inflammation. *Antibodies*. 2(3):392–414.

Stadlmann J, Pabst M, Kolarich D, Kunert R, Altmann F. 2008. Analysis of immunoglobulin glycosylation by LC-ESI-MS of glycopeptides and oligosaccharides. *PROTEOMICS*. 8(14):2858–2871. doi:10.1002/pmic.200700968. [accessed 2020 May 27]. https://onlinelibrary.wiley.com/doi/abs/10.1002/pmic.200700968.

Stöckmann H, Adamczyk B, Hayes J, Rudd PM. 2013. Automated, high-throughput IgG-antibody glycoprofiling platform. *Analytical Chemistry*. 85(18):8841–8849.

Stöckmann H, O'Flaherty R, Adamczyk B, Saldova R, Rudd PM. 2015. Automated, high-throughput serum glycoprofiling platform. *Integrative Biology*. 7(9):1026–1032.

Stone M. 1977. An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion. *Journal of the Royal Statistical Society: Series B (Methodological)*. 39(1):44–47. doi:10.1111/j.2517-6161.1977.tb01603.x.

Taparra K, Tran PT, Zachara NE. 2016. Hijacking the Hexosamine Biosynthetic Pathway to Promote EMT-Mediated Neoplastic Phenotypes. *Frontiers in Oncology*. 6:85.

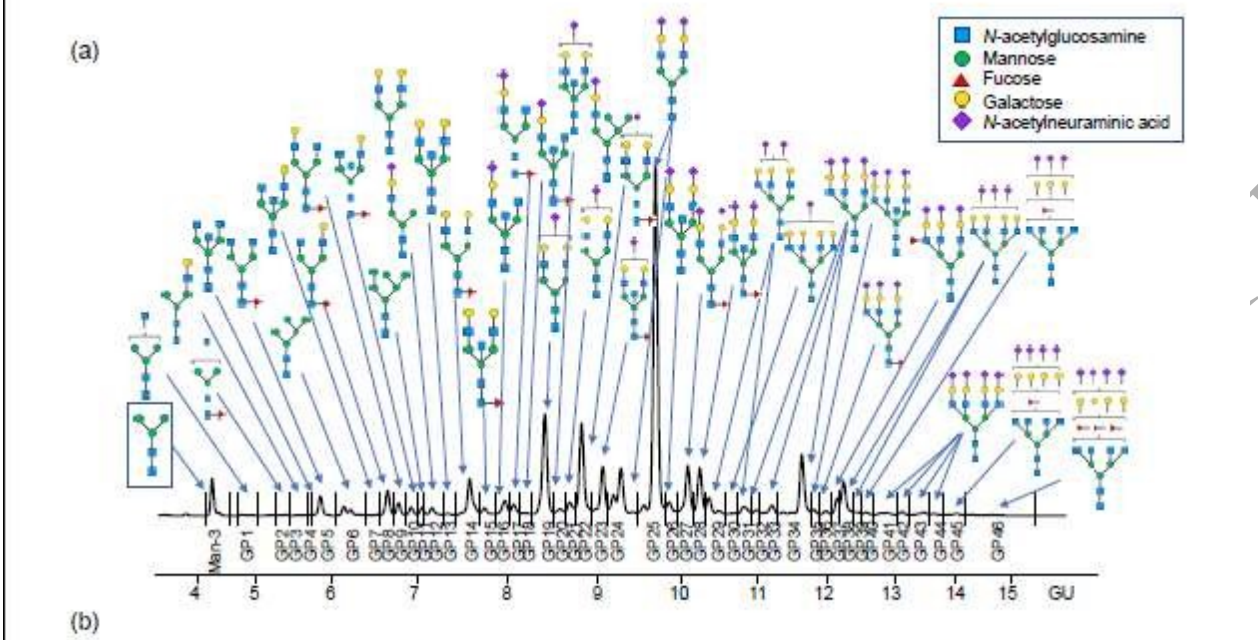
Yilmaz O, Afsar B, Ortiz A, Kanbay M. 2019. The role of endothelial glycocalyx in health and disease. *Clinical Kidney Journal*. 12(5):611–619.

Legends to figures

Figure 1

(a) Example of HILIC-UPLC chromatogram of *N*-glycans released from whole serum glycoproteins of diabetic patients. Each chromatogram was separated into 46 *N*-glycan peaks (GPs) containing one or more glycans in each GP according to Saldova et al. (Saldova et al. 2014), where all glycans are described in detail. Only the most abundant *N*-glycan structure in each peak is presented here. Blue squares represent *N*-acetylglucosamine, green circles represent mannose, red triangles represent fucose, yellow circles represent galactose and purple diamonds represent *N*-acetylneuraminic (sialic) acid.

(b) Summary of cross-sectional associations between each absolute quantitation of each GP and derived features with glycemic control (HbA_{1c}), ACR and eGFR at study day (models were adjusted for age, sex and diabetes duration); and prospective associations with incident albuminuria, final eGFR and incident final eGFR < 45 mL/min/1.73 m^2 (models were adjusted for age, sex, diabetes duration and length of follow-up; models for final eGFR and incident final eGFR were further adjusted for study day eGFR). The green and red labelled fields highlight the significant associations identified in this study. Positive associations are displayed in green and marked with a +, negative associations are shown in red and marked with a -. OligoM: Oligomannosylated, oaF: Outer arm fucosylated, CoreF: Core fucosylated, A1: Monoantennary, A2: Biantennary, A3: Triantennary, A4: Tetraantennary, A3A4: Tri- and tetra-antennary, G0: Agalactosylated, G1: Monogalactosylated, G2: Digalactosylated, G3: Trigalactosylated, G4: Tetragalactosylated, S0: Asialylated, S1: Monosialylated, S2: Disialylated, S3: Trisialylated and S4: Tetrasialylated structures.



(b)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
HbA_{1c}																+			+		+								+	+	+	+	+	+		+	+		+		+	+					
ACR												+				+			+						+	+			+		+	+		+		+	+	+	+		+	+					
eGFR					-	-		-	-	-	-			-			-		-		-				-	-	-		-		-			-		-	-	-	-			-	-	-	-
Inc. albuminuria																															+	+					-	-	-	-	+	+						
Final eGFR																		-		-						-			-	-	-	-	-	-			-	-		-		-		-			
Inc. eGFR < 45																															-	-						-		-		-					-

	OligoM	oaF	ooreF	A1	A2	A3	A4	A3A4	G0	G1	G2	G3	G4	S0	S1	S2	S3	S4
HbA_{1c}		+				+		+				+					+	
ACR		+				+	+	+				+	+			+	+	
eGFR	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Inc. albuminuria																		
Final eGFR		-				-		-				-					-	
Inc. eGFR < 45		-																

Table I: Participant characteristics at study day stratified by CKD stage

Covariate	G1 (N = 827, 52.8%)	G2 (N = 557, 35.6%)	G3 (N = 178, 11.4%)	All (N = 1565, 100%)	Missing
Main characteristics					
Age (years)	40.7 (29.0, 49.4)	54.9 (46.4, 63.4)	64.1 (55.5, 71.8)	47.8 (37.0, 58.6)	0
Sex (female)	41.0%	54.2%	56.2%	47.3%	0
Diabetes duration (years)	18.7 (9.9, 27.9)	25.9 (16.3, 36.1)	32.2 (21.0, 42.3)	22.2 (12.4, 32.7)	0
HbA_{1c} (mmol/mol)	70 (62, 83)	68 (60, 78)	68 (59, 79)	69 (61, 80)	3
HbA_{1c} (%)	8.6 (7.8, 9.7)	8.4 (7.6, 9.3)	8.4 (7.6, 9.4)	8.5 (7.7, 9.5)	3
Length of follow-up (years)	6.1 (5.3, 6.8)	6.4 (5.8, 7.0)	6.5 (5.5, 7.1)	6.3 (5.5, 6.9)	0
Kidney function					
Albumin/creatinine ratio (mg/mmol)	0.4 (0.2, 0.9)	0.5 (0.3, 1.4)	0.9 (0.4, 6.9)	0.4 (0.3, 1.1)	22
ACR category (normo/micro/macro)	88.5%/8.8%/2.5%	79.7%/12.6%/6.1%	65.7%/18.0%/14.6%	82.7%/11.2%/5.3%	13
Prevalent micro- or macro-albuminuria	11.5%	18.9%	33.1%	16.7%	0
Incident micro- or macro-albuminuria	14.6%	21.0%	26.4%	18.3%	0
eGFR (mL/min/1.73 m^2)	105.3 (97.3, 115.3)	74.1 (67.9, 81.5)	50.5 (42.1, 55.3)	91.7 (70.8, 106.0)	0
Prospective eGFR slope (mL/min/1.73 m^2/year)	-0.6 (-1.7, 0.6)	-0.6 (-2.2, 0.8)	-1.1 (-3.0, 0.5)	-0.6 (-2.0, 0.7)	101
Incident final eGFR < 45 mL/min/1.73 m^2	1.0%	5.9%	20.8%	5.0%	0
HDL-cholesterol (mmol/l)	1.5 (1.2, 1.8)	1.6 (1.3, 2.0)	1.5 (1.2, 1.9)	1.5 (1.2, 1.8)	58
Other covariates					
Total cholesterol (mmol/l)	4.6 (4.1, 5.2)	4.6 (4.0, 5.2)	4.2 (3.7, 4.9)	4.6 (4.0, 5.2)	23
Body mass index (kg/m^2)	26.0 (23.2, 29.6)	27.1 (24.5, 30.2)	27.4 (23.9, 31.1)	26.6 (23.8, 30.0)	8
Diastolic blood pressure (mmHg)	75.0 (69.0, 82.0)	75.0 (68.0, 81.0)	70.0 (63.0, 78.8)	75.0 (68.0, 81.0)	6
Systolic blood pressure (mmHg)	128.0 (119.0, 138.0)	134.0 (123.0, 146.0)	137.0 (123.2, 148.0)	130.0 (120.0, 141.0)	6
Ever smoker	62.6%	66.8%	71.9%	65.2%	0
On any anti-hypertensive treatment	28.9%	58.2%	86.5%	46.0%	0
On ACEi or ARB	27.0%	52.4%	78.7%	42.0%	0

We report frequency for categorical variables and median (IQR) for continuous variables
CKD stages are defined according to ranges of baseline eGFR in mL/min/1.73 m^2: G1: > 90; G2: 60–90; G3: 30–60
Not reported as a separate column are participants at stage G4 (eGFR: 15–30 mL/min/1.73 m^2; N = 3, 0.2%)

Table II: Top associations of study day HbA_{1c}, ACR and eGFR with *N*-glycan peaks and derived measures

		Absolute levels (ng/μL)		Percentage areas	
Glycan peak	**Main structures**	**β (95% CI)**	**p**	**β (95% CI)**	**p**
HbA_{1c}					
GP31	*A3G3S[3,3]2	0.19 (0.14, 0.24)	9.4×10^{-14}	0.27 (0.22, 0.32)	5.2×10^{-28}
GP37	A3F1G3S[3,3,3]3	0.15 (0.11, 0.20)	6.8×10^{-10}	0.18 (0.14, 0.23)	1.4×10^{-13}
GP36	FA3G3S[3,6,6]3	0.15 (0.10, 0.20)	1.4×10^{-09}	0.16 (0.12, 0.21)	4.4×10^{-11}
GP39	A4G4S[3,3,6]3 A4G4S[3,6,6]3	0.14 (0.09, 0.19)	9.8×10^{-09}	0.18 (0.13, 0.23)	1.5×10^{-12}
A3	Triantennary	0.14 (0.09, 0.19)	3.1×10^{-08}	0.18 (0.14, 0.23)	2.1×10^{-13}
G3	Trigalactosylated	0.14 (0.09, 0.19)	3.1×10^{-08}	0.18 (0.14, 0.23)	2.1×10^{-13}
OuterArmF	Outer arm fucosylated	0.13 (0.08, 0.18)	9.5×10^{-08}	0.16 (0.11, 0.21)	1.0×10^{-10}
GP34	A3G3S[3,3,6]3	0.13 (0.08, 0.18)	3.0×10^{-07}	0.13 (0.08, 0.18)	1.1×10^{-07}
S3	Trisialylated	0.13 (0.08, 0.18)	4.0×10^{-07}	0.17 (0.12, 0.22)	3.7×10^{-11}
GP29	A3G3S[3,6]2	0.13 (0.08, 0.18)	4.2×10^{-07}	0.16 (0.11, 0.21)	4.0×10^{-10}
ACR					
GP37	A3F1G3S[3,3,3]3	0.16 (0.11, 0.20)	8.4×10^{-10}	0.15 (0.10, 0.20)	8.8×10^{-09}
OuterArmF		0.14 (0.09, 0.19)	3.8×10^{-08}	0.13 (0.08, 0.18)	2.7×10^{-07}
GP31	*A3G3S[3,3]2	0.13 (0.08, 0.18)	1.1×10^{-07}	0.11 (0.06, 0.16)	1.0×10^{-05}
GP39	A4G4S[3,3,6]3 A4G4S[3,6,6]3	0.14 (0.09, 0.19)	1.1×10^{-07}	0.11 (0.06, 0.16)	7.6×10^{-06}
GP32	*A3G3S[3,3,3]3	0.14 (0.09, 0.19)	1.2×10^{-07}	0.09 (0.04, 0.14)	3.2×10^{-04}
S3	Trisialylated	0.12 (0.07, 0.17)	3.8×10^{-06}	0.09 (0.04, 0.14)	4.7×10^{-04}
A3	Triantennary	0.12 (0.07, 0.16)	5.6×10^{-06}	0.08 (0.03, 0.13)	1.7×10^{-03}
G3	Trigalactosylated	0.12 (0.07, 0.16)	5.6×10^{-06}	0.08 (0.03, 0.13)	1.7×10^{-03}
A3A4	Tri- and tetra-antennary	0.11 (0.07, 0.16)	6.7×10^{-06}	0.08 (0.03, 0.13)	1.1×10^{-03}
GP40	A4F1G3S[3,3,3]3	0.11 (0.06, 0.16)	1.7×10^{-05}	0.05 (0.00, 0.11)	3.3×10^{-02}
eGFR					
GP26	A2BG2S[3,6]2	-0.12 (-0.16, -0.09)	2.9×10^{-11}	-0.14 (-0.17, -0.10)	8.7×10^{-13}
GP32	*A3G3S[3,3,3]3	-0.11 (-0.14, -0.07)	1.8×10^{-08}	-0.06 (-0.09, -0.02)	2.2×10^{-03}
GP37	A3F1G3S[3,3,3]3	-0.09 (-0.13, -0.06)	2.4×10^{-07}	-0.05 (-0.08, -0.01)	1.1×10^{-02}
GP17	FA2[3]G1S[6]1	-0.10 (-0.13, -0.06)	2.5×10^{-07}	-0.05 (-0.08, -0.01)	7.4×10^{-03}
GP25	A2G2S[3,6]2	-0.09 (-0.13, -0.06)	3.3×10^{-07}	-0.02 (-0.05, 0.02)	3.4×10^{-01}
S0	Asialylated	-0.09 (-0.13, -0.06)	3.9×10^{-07}	-0.03 (-0.06, 0.01)	1.4×10^{-01}
A2	Biantennary	-0.09 (-0.13, -0.06)	4.6×10^{-07}	0.00 (-0.03, 0.04)	8.0×10^{-01}
OligoM	Oligomannosylated	-0.09 (-0.13, -0.06)	6.0×10^{-07}	-0.04 (-0.08, -0.00)	3.4×10^{-02}
G1	Monogalactosylated	-0.09 (-0.13, -0.06)	7.3×10^{-07}	-0.02 (-0.05, 0.02)	3.4×10^{-01}
OuterArmF	Outer arm fucosylated	-0.09 (-0.13, -0.06)	7.5×10^{-07}	-0.03 (-0.07, 0.00)	9.2×10^{-02}

Linear regression models were adjusted for age, sex and diabetes duration

Effect sizes are per unit of standard deviation of Gaussianized HbA_{1c} (HbA_{1c} section), and per unit of standard deviation of gaussianized glycan peak or derived measure (ACR and eGFR sections)

Bonferroni-corrected statistical significance threshold: $p < 3.9 \times 10^{-4}$

Only the predominant glycan structures are reported

Structure abbreviations: all *N*-glycans have two core GlcNAcs; F at the start of the abbreviation indicates a core fucose α1,6-linked to the inner GlcNAc; Mx, number (x) of mannose on core GlcNAcs; Ax, number of antenna (GlcNAc) on trimannosyl core; A2, biantennary with both GlcNAcs as β1,2-linked; A3, triantennary with a GlcNAc linked β1,2 to both mannose and the third GlcNAc linked β1,4 to the α1,3 linked mannose; A4, GlcNAcs linked as A3 with additional GlcNAc β1,6 linked to α1,6 mannose; B, bisecting GlcNAc linked β1,4 to β1,3 mannose; Gx, number (x) of β1,4 linked galactose on antenna; F(x), number (x) of fucose linked α1,3 to antenna GlcNAc; Sx, number (x) of sialic acids linked to galactose. Dx: isoforms with different mannose-binding. *sialic acids isomers (same composition but different sialic acid linkage arrangements resulting in different GUs from the original structures)

Table III: Top associations of absolute levels of *N*-glycan peaks and derived measures with incident albuminuria, final eGFR, and incident final eGFR < 45 mL/min/1.73 m^2

		Without adjustment for HbA_{1c}		Adjusted for HbA_{1c}	
Glycan peak	**Main structures**	**β / Odds ratio (95% CI)**	**p**	**β / Odds ratio (95% CI)**	**p**
Incident albuminuria					
GP32	*A3G3S[3,3,3]3	1.36 (1.19, 1.56)	1.2×10^{-05}	1.33 (1.15, 1.53)	7.7×10^{-05}
GP40	A4F1G3S[3,3,3]3 A4F1G3S[3,3,6]3 A4F1G3S[3,6,6]3	1.31 (1.14, 1.50)	1.7×10^{-04}	1.27 (1.11, 1.47)	8.3×10^{-04}
GP31	*A3G3S[3,3]2	1.29 (1.13, 1.48)	2.0×10^{-04}	1.23 (1.07, 1.41)	3.2×10^{-03}
GP39	A4G4S[3,3,6]3 A4G4S[3,6,6]3	1.28 (1.12, 1.47)	4.2×10^{-04}	1.23 (1.07, 1.42)	3.3×10^{-03}
OuterArmF	Outer arm fucosylated	1.28 (1.11, 1.47)	6.6×10^{-04}	1.23 (1.07, 1.42)	4.4×10^{-03}
GP38	A4G4S[3,3,3]3	1.27 (1.10, 1.45)	8.4×10^{-04}	1.25 (1.09, 1.44)	1.9×10^{-03}
GP41	A4G4S[3,3,3,3]4	1.26 (1.10, 1.45)	8.8×10^{-04}	1.24 (1.08, 1.42)	2.9×10^{-03}
A4	Tetraantennary	1.27 (1.10, 1.46)	1.2×10^{-03}	1.24 (1.08, 1.44)	3.2×10^{-03}
G4	Tetragalactosylated	1.27 (1.10, 1.46)	1.2×10^{-03}	1.24 (1.08, 1.44)	3.2×10^{-03}
GP12	A1[3]G1S[3]1 A2G2	1.25 (1.09, 1.43)	1.2×10^{-03}	1.24 (1.09, 1.43)	1.7×10^{-03}
Final eGFR					
GP37	A3F1G3S[3,3,3]3	-0.08 (-0.11, -0.05)	8.4×10^{-08}	-0.07 (-0.09, -0.04)	1.2×10^{-05}
GP31	*A3G3S[3,3]2	-0.08 (-0.10, -0.05)	2.5×10^{-07}	-0.06 (-0.09, -0.03)	5.5×10^{-05}
A3	Triantennary	-0.07 (-0.10, -0.04)	3.7×10^{-06}	-0.06 (-0.08, -0.03)	1.6×10^{-04}
G3	Trigalactosylated	-0.07 (-0.10, -0.04)	3.7×10^{-06}	-0.06 (-0.08, -0.03)	1.6×10^{-04}
OuterArmF	Outer arm fucosylated	-0.07 (-0.10, -0.04)	6.8×10^{-06}	-0.05 (-0.08, -0.03)	2.6×10^{-04}
GP34	A3G3S[3,3,6]3	-0.07 (-0.09, -0.04)	1.0×10^{-05}	-0.05 (-0.08, -0.03)	3.1×10^{-04}
GP29	A3G3S[3,6]2	-0.06 (-0.09, -0.03)	1.7×10^{-05}	-0.05 (-0.08, -0.02)	4.0×10^{-04}
A3A4	Tri- and tetra-antennary	-0.06 (-0.09, -0.03)	2.2×10^{-05}	-0.05 (-0.08, -0.02)	5.3×10^{-04}
S3	Trisialylated	-0.06 (-0.09, -0.03)	2.5×10^{-05}	-0.05 (-0.08, -0.02)	6.2×10^{-04}
GP32	*A3G3S[3,3,3]3	-0.06 (-0.09, -0.03)	5.1×10^{-05}	-0.05 (-0.08, -0.02)	9.6×10^{-04}
Incident final eGFR < 45 mL/min/1.73 m^2					
GP32	*A3G3S[3,3,3]3	1.56 (1.25, 1.95)	8.9×10^{-05}	1.45 (1.16, 1.82)	1.1×10^{-03}
GP31	*A3G3S[3,3]2	1.60 (1.26, 2.02)	8.9×10^{-05}	1.43 (1.12, 1.82)	3.7×10^{-03}
GP37	A3F1G3S[3,3,3]3	1.56 (1.24, 1.96)	1.1×10^{-04}	1.45 (1.14, 1.83)	2.0×10^{-03}
GP39	A4G4S[3,3,6]3 A4G4S[3,6,6]3	1.57 (1.24, 1.98)	1.3×10^{-04}	1.45 (1.15, 1.85)	2.0×10^{-03}
GP46	A4F3G4S[3,3,3,3]4	1.53 (1.22, 1.93)	2.2×10^{-04}	1.51 (1.20, 1.93)	6.3×10^{-04}
OuterArmF	Outer arm fucosylated	1.66 (1.27, 2.18)	2.6×10^{-04}	1.52 (1.16, 2.01)	2.6×10^{-03}
GP41	A4G4S[3,3,3,3]4	1.53 (1.21, 1.93)	3.0×10^{-04}	1.45 (1.15, 1.84)	1.9×10^{-03}

GP29	A3G3S[3,6]2	1.52 (1.19, 1.94)	8.4×10^{-04}	1.42 (1.11, 1.82)	5.6×10^{-03}
GP40	A4F1G3S[3,3,3]3 A4F1G3S[3,3,6]3 A4F1G3S[3,6,6]3	1.48 (1.17, 1.86)	8.5×10^{-04}	1.39 (1.10, 1.75)	5.8×10^{-03}
GP21	M5A1G1S[3]1 *A2G2S[3]1	1.51 (1.17, 1.93)	1.2×10^{-03}	1.44 (1.12, 1.85)	3.8×10^{-03}

Logistic regression models for incident albuminuria were adjusted for age, sex, diabetes duration and length of follow-up; linear regression models for final eGFR and logistic regression models for incident final eGFR < 45 mL/min/1.73 m^2 were adjusted for age, sex, diabetes duration, study day eGFR and length of follow-up; column headers indicate whether each model was further adjusted for HbA_{1c}

Effect sizes are per unit of standard deviation of Gaussianized glycan peak and derived measure

Bonferroni-corrected statistical significance threshold: $p<3.9\times10^{-4}$

Only the predominant glycan structures are reported.

Structure abbreviations are in Table II.