**IgG glycan patterns are associated with type 2 diabetes in independent European populations**

Roosmarijn F.H. Lemmersa,b, \*Marija Vilajc, \*Daniel Urdad, Felix Agakovd, Mirna Šimurinae, Lucija Klaricc,e,f, Igor Rudang, Harry Campbellg, Caroline Haywardf, Jim F. Wilsonf,g , Aloysius G. Lieverseb, Olga Gornike, Eric J.G. Sijbrandsa, Gordan Laucc,e, Mandy van Hoeka.

\*These authors contributed equally

a Department of Internal Medicine, Erasmus MC - University Medical Center Rotterdam, the Netherlands

b Department of Internal Medicine, Maxima Medical Center, Eindhoven, The Netherlands

c Genos Glycoscience Research Laboratory, Zagreb, Croatia

d Pharmatics Limited, Edinburgh, United Kingdom

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

f MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom

g Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom

**Corresponding author**

Mandy van Hoek

Erasmus MC- room D432, PO-box 2040, 3000CA,

Rotterdam, The Netherlands

Tel +31 633343603,

Email: m.vanhoek@erasmusmc.nl

Number of words (text): 4072. Number of words abstract: 241

Number of tables: 5, Number of Figures: 4, Number of supplements: 7.

**Abstract**

**Background;** Type 2 diabetes results from interplay between genetic and acquired factors. Glycans on proteins reflect genetic, metabolic and environmental factors. However, associations of IgG glycans with type 2 diabetes have not been described. We compared IgG N-glycan patterns in type 2 diabetes with healthy subjects.

**Methods;** In the DiaGene study, a population-based case-control study, (1886 cases and 854 controls) 58 IgG glycan traits were analyzed. Findings were replicated in the population-based CROATIA-Korcula-CROATIA-Vis-ORCADES studies (162 cases and 3162 controls), and meta-analyzed. AUCs of ROC-curves were calculated using 10-fold cross-validation for clinical characteristics, IgG glycans and their combination.

**Results;** After correction for extensive clinical covariates, 5 IgG glycans and 13 derived traits significantly associated with type 2 diabetes in meta-analysis (after Bonferroni correction). Adding IgG glycans to age and sex increased the AUC from 0.542 to 0.734. Adding them to the extensive model did not substantially improve the AUC. The AUC for IgG glycans alone was 0.729.

**Conclusions;** Several IgG glycans and traits firmly associate with type 2 diabetes, reflecting a pro-inflammatory and biologically-aged state. IgG glycans showed limited improvement of AUCs. However, IgG glycans showed good prediction alone, indicating they may capture information of combined covariates. The associations found may yield insights in type 2 diabetes pathophysiology.

**General significance;** This work shows that IgG glycomic changes have biomarker potential and may yield important insights into pathophysiology of complex public health diseases, illustrated here for the first time in type 2 diabetes.

**Key words:** IgG, glycosylation, type 2 diabetes, prediction, inflammation, ageing

**Abbreviations**

American Diabetes Association (ADA)

Area under the curve (AUC)

Chronic Kidney Disease (CKD)

Fucosylated diantennary glycans (FA2)

Genome wide association studies (GWAS)

Glycan Peak (GP)

Hydrophilic interaction liquid chromatography solid phase extraction (HILIC-SPE)

Inflammatory Bowel Disease (IBD),

Rheumatoid Arthritis (RA)

Receiver Operator characteristic (ROC)

Systemic Lupus Erythematosus (SLE)

Ultra-performance liquid chromatography (UPLC)

World Health Organization (WHO)

**Introduction**

Type 2 diabetes is an extremely challenging health issue in the 21st century [1]. It is a multifactorial disease, resulting from intricate interplay between environmental and genetic factors. Although many environmental and genetic risk factors have been identified, the underlying mechanisms of the disease remain largely unknown.

Genetic factors identified so far explain up to 5-10% of disease risk [2]. It is likely that epigenetic and posttranscriptional modifications play a substantial role in disease pathophysiology. Glycosylation of proteins is one of the most common posttranslational modifications. *N-*glycans are present on most proteins and are generated through interplays between hundreds of enzymes [3, 4]. They play structural, functional, and regulatory roles and reflect genetic, metabolic, and environmental factors [3]. These unique qualities raises interest for the role and use of *N-*glycans in complex disease such as type 2 diabetes. Total *N*-glycomic changes have been reported once in a relatively large type 2 diabetes population and showed significant changes. However, total *N-*glycomic changes do not inform us on the exact underlying proteins and processes involved and results are influenced by the relative abundance of proteins in the circulation.

Type 2 diabetes is characterized by a pro-inflammatory state and elevated levels of inflammatory markers, such as C-reactive protein and interleukin-6, that have been associated with risk of developing type 2 diabetes [5]. IgG *N-*glycome changes have been linked to clinical risk factors for type 2 diabetes, such as age, BMI, smoking, and dyslipidemia [6-9]. IgG is a tetramer protein complex consisting of Fc and Fab regions. IgG glycosylation is particularly interesting, as *N-*glycans attached to the Fc part of IgG can modulate and switch its function from pro- to anti-inflammatory and vice versa [3, 10, 11]. However, IgG *N-*glycosylation patterns in type 2 diabetes compared to the non-diabetic state have not been described in current literature. Knowledge on specific IgG *N-*glycosylation profiles in type 2 diabetes could shed light on underlying inflammatory pathophysiological processes and on drug target and biomarker potential.

We hypothesized that characteristic IgG *N-*glycan profiles are present in type 2 diabetes. In this study, we investigated the association of IgG *N-*glycan profiles with type 2 diabetes in a large population-based case control cohort, followed by replication in independent European samples. Moreover, we evaluate the AUC under the Receiver Operator Characteristic (ROC) curves for models including clinical characteristics, IgG glycans, and both.

**Research design and Methods**

*Subjects*

Discovery cohort:

The discovery cohort was the DiaGene study. Characteristics of the DiaGene population have been described elsewhere [12]. Briefly, the DiaGene study is a case-control cohort collected in and around the city of Eindhoven, the Netherlands. All hospitals in this area participated, as well as the center for primary care diagnostics. Type 2 diabetes was diagnosed according the ADA and WHO criteria [13, 14]. Patients with other types of diabetes, diabetes secondary to Cushing’s syndrome or to corticosteroid use were excluded. Controls were recruited by advertising in local newspapers (aged 55 years or older) and through unrelated friends/family of the cases. Controls with diabetes mellitus, impaired glucose tolerance or Cushing’s disease were excluded. In total, 1886 patients and 854 controls were included. All participants gave written informed consent. The study has been approved by the Medical Ethical Committee of the Erasmus MC.

Replication cohort:

We validated the results of our discovery analyses in three family-based cohorts from isolated islands: the Orkney Complex Disease Study (ORCADES), CROATIA-Vis, and CROATIA-Korcula. The ORCADES study was collected between 2005 and 2011 from the Orkney isles in Scotland. The CROATIA-Vis and CROATIA-Korcula cohorts were collected from the Croatian islands Vis and Korcula in 2003-2004 and 2007, respectively. Type 2 diabetes status was self-reported. These three deeply phenotyped cohorts have been described in more detail elsewhere [15-18]. Procedures used for extraction and preprocessing of glycomic profiles were harmonized for the discovery and replication cohorts. After quality control, the total number of samples in the combined replication cohorts was n=3324 (162 cases with type 2 diabetes and 3162 controls).

*IgG N-glycosylation*

In the DiaGene study, 1837 cases and 852 controls had plasma available for IgG glycosylation analyses. Twenty nine samples failed quality control. The remaining 1815 cases and 845 controls were included in the analyses. IgG glycan isolation, release and labeling in the DiaGene study were executed as described previously [19]. In total, 24 IgG glycan peaks were measured. The same peaks were measured in the CROATIA-Vis, CROATIA-Korcula and ORCADES samples as more extensively described by Pucic et al [19]. Figure 1 shows the glycan structures of the most abundant glycans per peak; a more detailed description can be found in supplementary table S1.

In short, for both the discovery and the replication cohorts, IgG was isolated from plasma using 96-well protein G monolithic plates, eluted with 0.1M formic acid, and neutralized with 1M ammonium bicarbonate. Dried IgG samples (150-200ug) were denatured (with 30μl SDS), incubated at 65°C for 10 minutes and cooled to room temperature, followed by addition of 10μl of Igepal-CA630 (φ=4%). *N-*glycans were released with the addition of 1.2U of PNGase F in 10μL 5x PBS and incubation at 37°C for 18 hours. Released *N-*glycans were labeled with 2-AB. The samples were cleaned by using hydrophilic interaction liquid chromatography solid phase extraction (HILIC-SPE) and stored at -20°C until ultra-performance liquid chromatography (UPLC). Finally, 24 IgG glycan peaks (GPs) were measured by Waters Acquity UPLC instrument as described previously [19]. All chromatograms were separated in the same manner into 24 peaks and the amount of glycans in each peak was expressed as percentage of total integrated area.

An additional 34 derived IgG glycan traits were calculated from the 24 directly measured peaks as described previously [6, 19]. These traits represent specific IgG glycosylation features (such as galactosylation, fucosylation and sialylation); a detailed description can be found in supplementary table S1.

*Harmonization*

Twelve glycan profiles were removed after visual inspection of the chromatograms. In all cohorts, preprocessing of the data followed the same protocol, where: (i) IgG glycan expressions were normalized by the total area under the expression peaks, followed by log-transformation, and (ii) batch effects were removed by applying ComBat modified to correct for outliers (using R Package sva). Measurement error of each of the 24 glycan structures was estimated based on the correlation between replicated samples (n=69 out of 2688 samples) and variation of standard sample measurements (n=153), which are shown in supplementary figure S1 and S2, respectively. No samples were excluded based on measurement error. After the preprocessing, ORCADES, CROATIA-Vis and CROATIA-Korcula cohorts were pooled, to form a joint validation dataset.

*Statistical analyses*

The distribution of the IgG glycan peaks was analyzed by visual inspection of QQ plots and showed no major deviations from normality. Pairwise dependencies between the 24 IgG glycan peaks and age, BMI, cholesterol traits, and creatinine in the discovery population were evaluated by Pearson’s correlation coefficients; significance was assessed by the correlation test for bivariate normal distributions. Associations of smoking and sex with the IgG glycan peaks were evaluated with ANOVA. Non-HDL cholesterol was calculated as the difference between total cholesterol and HDL-cholesterol. These clinical variables were selected because they are risk factors for type 2 diabetes [20] and/or associated with N-glycosylation [6, 8, 21]. Student’s t-tests and chi-squared tests were applied to test differences between cases and controls for continuous and categorical variables respectively.

Logistic regression models were used to investigate the associations of directly measured IgG glycan peaks and derived traits with type 2 diabetes. Three models were considered: model 1) adjusted for age and sex; model 2) adjusted for age, sex, and BMI; and model 3) adjusted for age, sex, BMI, HDL-cholesterol, non-HDL-cholesterol, smoking and creatinine. All analyses described above were performed in the discovery population using ‘binary logistic regression’ in SPSS Statistics, version 21. Covariates were entered without forward or backward selection.

All analyses in the validation cohort were done in R, using the packages glm, pROC, and ROCR [22]. The likelihood-ratio test was used to compare the performance of logistic regression models with and without IgG glycans. The same clinical and demographic variables were applied in the models of the validation cohort. All analyses in the validation cohort were additionally adjusted for the covariate denoting their original cohort.

Results of the association analyses in the discovery and replication sets were meta-analyzed using the weighted z-transform method [23]. This method takes into account the regression coefficients, standard errors, and number of participants per study, and sets the weight of each study to the square root of the sample size. Bonferroni correction for multiple testing was applied for 58 tests (24 direct and 34 derived measurements). P-values < 8.62E-4 were considered statistically significant.

For prediction analyses, we fitted logistic regression models using 10-fold cross-validation. In this method, the original sample is partitioned in 10 non-overlapping subsamples of equal size, where 1 sample serves as a test set and the other 9 samples serve as training sets, and models are fitted repeatedly for all training/test fold partitions. We first computed the average AUC across the test folds, and finally the aggregated AUC, where predictions on 10 independent subsets were merged into a single vector. This method limits the danger of overfitting and estimates how well the model can be generalized to independent datasets. We compared model AUCs that included the above-mentioned clinical variables with those that additionally contained the 24 directly measured IgG glycans. All continuous variables were standardized before they were included in the models for AUC calculation.

The term prediction is used in the statistical sense, *i.e.* the inference of an outcome given the covariates. Specifically, we do not refer to forecasting of prospective outcomes, and our predictions in the considered case-control setting refer to the time of the measurements.

**Results**

*Cohort characteristics*

Characteristics of the cohorts are shown in Table 2. In the DiaGene study, patients with type 2 diabetes were more often male (53.7 vs 40.2% for cases and controls, respectively), had higher BMI and creatinine levels, lower HDL-cholesterol and non-HDL cholesterol levels, and were more often smokers than healthy controls. In the replication cohort, cases were older (66.6 vs. 54.5y), had a lower HDL-cholesterol, and higher BMI and creatinine.

*Clinical traits associated with type 2 diabetes or N-glycosylation*

Correlations of clinical traits associated with type 2 diabetes or N-glycosylation with the 24 IgG glycan peaks are shown in table 2. IgG glycan peaks show significant univariate associations with more than one clinical trait and vice versa.

*IgG glycan associations with type 2 diabetes*

Illustrative IgG glycan profiles for type 2 diabetes and the non-diabetic state are shown in figure 2; median and interquartile ranges of all peaks in both cohorts are shown in supplementary table S2. Table 3 shows glycans that were significantly associated with type 2 diabetes after meta-analyses in the discovery and validation cohorts, adjusted for age and sex. The strongest associations are for glycan peaks GP6, GP8 and GP9 (P meta-analyses <2.22E-16). Additional adjustment for BMI did not substantially change significant associations (complete data on all outcomes in all models are shown in Supplementary tables S3-S5).

The associations with type 2 diabetes remained statistically significant for glycan peaks GP6, GP8, GP9, GP10, and GP11 in the full model (adjusted for age, sex, BMI, HDL-cholesterol, non-HDL-cholesterol, smoking and creatinine). Figure 3 shows boxplots for these IgG glycan peaks; β- and p-values adjusted for the full model are shown in table 4. Excluding participants using medication indicative of auto-immune, malignant, or inflammatory conditions from the analyses did not change our results (data not shown). These glycans are reflective of increased presence of agalactosylated (GP6) and monogalactosylated (GP10, GP11) FA2 glycans with bisecting *N-*Acetylglucosamine (GlcNAc) and decreased presence of FA2[6] (GP8) and FA2[3] (GP9) monogalactosylated glycans. The derived traits associated with type 2 diabetes show a decrease in the percentage of sialylation of all fucosylated structures with bisecting GlcNAc (FBGS/(FB+FBG+FBGS)) and an increase of bisecting GlcNAc in fucosylated disialylated structures (FBS2/FS2 and FBS2/(FS2+FBS2)). There was a decrease in monogalactosylated structures (G1n). Furthermore, there was a decrease in fucosylated structures without bisecting GlcNAc in neutral glycans (Fn) and in agalactosylated (FG0n/G0n) and monogalactosylated (FG1n/G1n) structures. Finally, there was an increase in fucosylated structures with bisecting GlcNAc (FBn, FBG0n/G0n, FBG1n/G1n, FBG2n/G2n, FBn/Fn, FBn/Fntotal).

*Area under the ROC curves for type 2 diabetes*

The average AUCs for 24 IgG glycans without clinical variables was 0.73 and 0.75 for the discovery and validation populations, respectively. Figure 4 shows ROC curves for the predictive capacity of clinical variables with and without glycans for type 2 diabetes in the case-control setting.

As shown in table 5, adding the 24 IgG glycans to clinical variables led to improvement of the AUC in all models. This improvement was most extensive in the model containing age and sex (AUC 0.542 vs 0.734 for age and sex vs age, sex, and 24 IgG glycans). Similar results were found in our validation cohort. Augmented AUCs were similar and differences with average AUCs were negligible (data not shown).

**Discussion**

In the present study, we are the first to describe the association and predictive potential of IgG *N-*glycans in type 2 diabetes versus healthy individuals in a large case-control study. We have found significant and replicated associations for GP6, GP8, GP9, GP10, and GP11 and several derived IgG glycan traits associated with decreased galactosylation and sialyation, an increase in fucosylated structures with bisecting GlcNAc and a decrease in fucosylated structures without bisecting GlcNAc. AUCs showed slight improvement for predicting type 2 diabetes in the case-control setting when IgG glycan peaks were added to models with clinical characteristics. Remarkably, the AUC for IgG glycans by themselves for type 2 diabetes was 0.73.

Type 2 diabetes is a multifactorial disorder in which genetics, environmental, and metabolic influences interact at multiple levels. Glycans reflect all these influences, which makes them attractive biomarkers for multifactorial diseases. In this study, the IgG glycan profile showed significant associations with age, sex, BMI, lipid profile, smoking, and kidney function. These associations are largely in line with previous reports on IgG glycosylation and total *N-*glycome in these traits [6-8, 21, 24].

The associations we found in type 2 diabetes are reflective of an overall pro-inflammatory state and biological aging [3, 7, 11, 25-28]. In the most extensive model, the associations were in concordance with a decrease in galactosylation and sialylation, an increase in fucosylated structures with bisecting GlcNac, and a decrease in fucosylated structures without bisecting GlcNAc.

Galactosylation has been shown to influence the inflammatory potential of IgG [25]. Lack of galactosylation is associated with a pro-inflammatory state of IgG through activation of the complement cascade [27]. Previous studies found a decrease in galactosylation in rheumatoid arthritis (RA) [29, 30], systemic lupus erythematosus (SLE) [31], inflammatory bowel disease (IBD) [32]), chronic kidney disease (CKD) [21], hypertension [9], colorectal cancer [33], and with increasing age [7] and BMI [6]. Contrarily, in Parkinson’s disease, galactosylation increased [34]. The decreased IgG galactosylation found in most studies followed the same patterns as in type 2 diabetes and thus seems a general feature of pro-inflammatory disease. A possible explanation for this decreased galactosylation are posttranslational modifications of the enzyme β4-galactosyltransferase-1 [35]. Interestingly, Ercan et al. recently showed that estrogen increases galactosylation in IgG and accounts for a difference between men and premenopausal women [24]. As the majority of individuals in our discovery and replication cohorts were older than 50 and all models were adjusted for sex, this is not expected to influence the findings in our study.

Sialylation also plays an important role in the inflammatory potential of IgG. Addition of sialic acid to IgG converts its function from pro- to anti-inflammatory by decreased binding to Fcγ receptors [11]. It should be noted that a lack of galactose also leads to a lack of sialic acid, as sialic acid mainly binds to terminal galactose [35]. Similar to our findings, a reduction in sialylation was recently described in SLE [31], IBD [32], CKD [21], and colorectal cancer [33], while sialylation increased in hypertension [9]. Interestingly, in our study, decreased sialylation was present in fucosylated bisecting structures mainly due to decreased GP19 (FA2BG2S1). In contrast, in SLE and IBD, sialylation decreased in general while GP19 increased [31, 32]; in CKD, sialylation was decreased in fucosylated non-bisecting structures due to a decrease in GP18 (FA2G2S1; analysis adjusted for diabetes). Decreased sialylation was seen in roughly the same glycan peaks in individuals with colorectal cancer patients as in our study, although different derived traits were calculated. Thus, although decreased IgG sialylation is seen in more diseases than type 2 diabetes, the desialylated glycan structures differ and could thus be disease-specific changes.

Finally, core-fucosylation and bisecting GlcNAc exert important effects on IgG function. Core fucosylation prevents antibody-dependent cytotoxicity [26, 28, 36], while the presence of bisecting GlcNAc is thought to have the opposite effect [37]. We observed decreased fucosylation in structures without bisecting GlcNAc and increased fucosylation in structures with bisecting GlcNAc*.* Moreover, we found increased bisection in fucosylated glycans, especially when sialylated. These changes indicate a higher antibody dependent cytotoxic potential of IgG in type 2 diabetes. Similar associations have been found in increasing age, CKD [21], SLE.[31], colorectal cancer [33], and hypertension [9], although small differences existed. In CKD, the strongest association with bisection was in fucosylated monosialylated structures, while in type 2 diabetes this was the case for fucosylated disialylated structures. Furthermore, bisection of neutral structures was positively associated with type 2 diabetes, while it was negatively associated with colorectal cancer [33] and not at all with CKD [21]. If and how these differences in bisection and fucosylation affect the efficacy of IgG in antibody-dependent cytotoxicity is unknown and should be subject of future studies. A potential gene of interest for studying this is *MGAT-3*, which encodes N-acetylglycosaminyltransferase 3, a glycosyltransferase that adds bisecting GlcNAc to IgG glycans, and is associated with IgG bisecting [38].

Recently, Testa et al described the total *N-*glycome in type 2 diabetes [39] and found a reduction in monogalactosylation on the 3- and 6-arm of fucosylated diantennary glycans, which is in line with our findings. In contrast with our findings, Itoh et al found [40] a small increase in an *N-*glycan with α1,6-core fucose in the total *N-*glycome in type 2 diabetes. Findings of both studies are difficult to compare to ours, due to the different quantification techniques for *N-*glycans. The method of Testa et al. does not allow for measurement of sialylation and has lower resolution. Itoh et al. measured exact bi-antennary glycan structures as single molecules, while our method divides IgG glycans in groups, reflecting major structural characteristics (such as total core fucosylation, outer arm fucosylation, degree of sialylation, etc.).In both studies, total *N-*glycome was measured, quantifying glycans from all plasma proteins, while we quantified IgG-specific glycans. Total plasma N-glycome not only reflects *N-*glycosylation, but also the relative abundance of proteins in the circulation.

Due to the case-control nature of our study it remains unclear whether the IgG glycan associations are a cause or a consequence of the disease. Notably, for RA it was shown that a decrease in galactosylation precedes a flare of the disease rather than the other way around [41]. In addition, the IgG glycans associated with type 2 diabetes in our study showed high heritabilities in previous studies (35-76%) [19, 42]. Genome wide association studies (GWAS) have identified several loci influencing *N-*glycosylation. Of interest, the *HNF1A* gene, known for its causal role in monogenic types of diabetes and its common variants associated with type 2 diabetes [12], is a genetic locus of significant association with total plasma *N-*glycome as a regulator of fucosylation [43]. Moreover, *HNF1A* regulates expression of *FUT8* [43], a fucosyltransferase gene associated with IgG core fucosylation [38], which was associated with type 2 diabetes in our study. In addition, GWAS on IgG glycosylation found associations with the *ST6GAL1* gene [38], which was previously associated with type 2 diabetes in South Asians [44]. *ST6GAL1*, as well as *B4GALT1 (*encoding sialyltransferase 6 and beta1,4-galactosyltransferase 1, respectively), are associated with IgG sialylation [38] and could therefore be candidate genes for future studies on the role of IgG sialylation in type 2 diabetes pathophysiology. Besides these genetic links between glycosylation and type 2 diabetes, the associations found between risk factors for type 2 diabetes and IgG glycans support the hypothesis that pro-inflammatory IgG could already at an early stage contribute to pathophysiology. Possibly, pro-inflammatory IgG could contribute to the pro-inflammatory state seen in obesity and type 2 diabetes [45], which leads to β-cell destruction and insulin resistance [46, 47]. A potential mechanism is through complement activation, which has been shown to be associated with type 2 diabetes incidence and complications [48, 49] and is activated by agalactosylated IgG.Taken together, these findings are supportive of the hypothesis that IgG glycans play a role in type 2 diabetes disease pathophysiology, rather than just reflective of changes the disease brought on.

As glycan profiles are associated with genetic, metabolic and environmental influences, this adds to their predictive potential. We observed a modest improvement when adding the IgG glycan profiles to clinical risk factors. However, the AUC of IgG glycans alone was substantial compared to combined clinical risk factors, indicating that the IgG glycan profile captures much of the combined risk of these factors. This underlines its potential as a biomarker of the complex inflammatory pathophysiological changes in type 2 diabetes. Future studies in prospective cohorts should be conducted to investigate whether this biomarker potential could be used for personalized approaches in prevention and treatment of the disease as well as its role in diabetes complications. Of note, because of the case-control setting, the predictive capacity of clinical risk factors was high in both cohorts, especially in the replication cohort where the cases were considerably older than the controls. The discovery population was quite well-matched for age. A consideration in the interpretation of our findings is that we have only looked at the glycan profile of IgG. As many more glycoproteins are present in the circulation, the total *N-*glycome or a combination of several glycoprotein profiles may give better results for prediction of type 2 diabetes.

Strengths of our study are the robust manner in which IgG *N-*glycans were measured, the large size of our populations, and the fact that we included an independent replication cohort. Also, we have calculated our AUC using 10-fold cross validation to limit overfitting and get an accurate idea on how the models will perform in independent datasets. A limitation is the lack of prospective follow-up for the outcome type 2 diabetes. Therefore, no causal claims can be made and AUCs are to be interpreted in a case-control setting of patients already treated for their disease. The case-control setting in general may lead to overestimation, while the fact that patients are being treated for their disease may reduce the contrast between the groups. The current setup does allow comparing how IgG glycans add to the models and how important their relative contribution is on their own, but does not give an absolute reflection of their predictive potential. Second, simple predictive models (logistic regression) were used. Residual confounding by yet unknown factors involved in type 2 diabetes pathophysiology and the IgG glycome cannot be excluded. Third, we cannot exclude an effect of glucose-lowering treatment on the associations between IgG glycans and type 2 diabetes. In the discovery cohort, 19.2% was treated with diet alone, 64.9% took oral glucose-lowering agents, and 31.6 used insulin (alone or in combination with oral medication); we do not have this data for the replication cohort. Collaborations to create larger sample sizes with glycomic and treatment information, preferably with a prospective setup, will enable us to research these effects in future. Fourth, the method used to measure IgG glycans does not allow to distinguish between Fc and Fab associated glycans. As only 15%-20% of the Fab part of IgG is glycosylated, the associations are more likely to be driven by differences in Fc glycans [50]. However, a contribution of changes in Fab N-glycosylation cannot be excluded.

To conclude, we found robust associations of IgG glycan peaks and traits with the presence of type 2 diabetes that reflect a pro-inflammatory and biologically aged state. Future studies should be directed at in-depth pathophysiological insights that can be derived from IgG glycan associations with type 2 diabetes. Also, prospective follow-up studies and genetic studies can shed light on the causality, its potential as a biomarker for complex inflammatory processes, and true predictive capacity of IgG and other glycans in type 2 diabetes.

**Acknowledgements**

**Funding**

MH was supported by the Erasmus MC fellowship. Glycan analysis was supported in part by FP7 projects HighGlycan (contract #278535), MIMOmics (contract #305280), MLPM (contract #316861), HTP-GlycoMet (contract #324400) and RegPot Integra-Life (contract #315997) from the European Commission and the HRZZ grant #7769.

**Duality of interest**

RL, MS, IR, HC, CH, JW, AL, OG, ES, MH declare no conflicts of interest

FA is a founder and co-owner of Pharmatics Limited. Pharmatics declares no conflicts of interest regarding this work.

DU is an employee of Pharmatics Ltd and has no conflicts of interest regarding this work.

GL is the founder and owner of Genos Ltd, which offers commercial service of glycomic analysis and has several patents in this field.

MV and LK are employees of Genos Ltd.

**Contribution statement**

RL database collection and maintenance DiaGene study database, wrote and reviewed/edited manuscript, approved the final manuscript

MV laboratory analyses of IgG glycans, computational processing of glycan data, reviewed/edited manuscript, approved the final manuscript

DU statistical analyses replication cohorts, k-fold AUC analysis, meta-analyses, reviewed/edited manuscript, approved the final manuscript

FA statistical analyses replication cohorts, k-fold AUC analysis, meta-analyses, reviewed/edited manuscript, approved the final manuscript

MŠ laboratory analyses of IgG glycans, computational processing of glycan data, reviewed/edited manuscript, approved the final manuscript

LK laboratory analyses of IgG glycans, computational processing of glycan data, reviewed/edited manuscript, approved the final manuscript

IR collection, design and coordination of CROATIA-Vis and CROATIA-Korcula, reviewed/edited manuscript, approved the final manuscript

HC collection, design and coordination of ORCADES, reviewed/edited manuscript, approved the final manuscript

CH collection and design of CROATIA and ORCADES studies, reviewed/edited manuscript, approved the final manuscript

JW collection, design, coordination of the ORCADES study, reviewed/edited manuscript, approved the final manuscript

AL collection, design and coordination of the DiaGene study, reviewed/edited manuscript, approved the final manuscript

OG laboratory analyses of IgG glycans, computational processing of glycan data, reviewed/edited manuscript, approved the final manuscript

ES conception of research question, design and coordination of the DiaGene study, reviewed/edited manuscript, approved the final manuscript

GL conception of research question, laboratory analyses of IgG glycans, reviewed/edited manuscript, approved the final manuscript

MH conception of the research question, wrote manuscript, collecting and coordination of the DiaGene study, statistical analyses, reviewed/edited manuscript, approved the final manuscript

References

[1] WHO (2016) Diabetes mellitus. Fact sheet N°138

[2] Morris AP, Voight BF, Teslovich TM, et al. (2012) Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 44: 981-990

[3] Gornik O, Pavic T, Lauc G (2012) Alternative glycosylation modulates function of IgG and other proteins - implications on evolution and disease. Biochim Biophys Acta 1820: 1318-1326

[4] Lauc G, Rudan I, Campbell H, Rudd PM (2010) Complex genetic regulation of protein glycosylation. Mol Biosyst 6: 329-335

[5] Wang X, Bao W, Liu J, et al. (2013) Inflammatory markers and risk of type 2 diabetes: a systematic review and meta-analysis. Diabetes Care 36: 166-175

[6] Nikolac Perkovic M, Pucic Bakovic M, Kristic J, et al. (2014) The association between galactosylation of immunoglobulin G and body mass index. Prog Neuropsychopharmacol Biol Psychiatry 48: 20-25

[7] Kristic J, Vuckovic F, Menni C, et al. (2014) Glycans are a novel biomarker of chronological and biological ages. J Gerontol A Biol Sci Med Sci 69: 779-789

[8] Knezevic A, Gornik O, Polasek O, et al. (2010) Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma N-glycans. Glycobiology 20: 959-969

[9] Wang Y, Klaric L, Yu X, et al. (2016) The Association Between Glycosylation of Immunoglobulin G and Hypertension: A Multiple Ethnic Cross-Sectional Study. Medicine (Baltimore) 95: e3379

[10] Anthony RM, Nimmerjahn F, Ashline DJ, Reinhold VN, Paulson JC, Ravetch JV (2008) Recapitulation of IVIG anti-inflammatory activity with a recombinant IgG Fc. Science 320: 373-376

[11] Kaneko Y, Nimmerjahn F, Ravetch JV (2006) Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. Science 313: 670-673

[12] Voight BF, Scott LJ, Steinthorsdottir V, et al. (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 42: 579-589

[13] Alberti KG, Zimmet PZ (1998) Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. Diabet Med 15: 539-553

[14] American Diabetes A (2011) Diagnosis and classification of diabetes mellitus. Diabetes Care 34 Suppl 1: S62-69

[15] McQuillan R, Leutenegger AL, Abdel-Rahman R, et al. (2008) Runs of homozygosity in European populations. Am J Hum Genet 83: 359-372

[16] Rudan I, Campbell H, Rudan P (1999) Genetic epidemiological studies of eastern Adriatic Island isolates, Croatia: objective and strategies. Coll Antropol 23: 531-546

[17] Polašek O, Marušić A, Rotim K, et al. (2009) Genome-wide Association Study of Anthropometric Traits in Korčula Island, Croatia. Croatian Medical Journal 50: 7-16

[18] Vitart V, Rudan I, Hayward C, et al. (2008) SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout 40: 437-442

[19] Pucic M, Knezevic A, Vidic J, et al. (2011) High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human populations. Mol Cell Proteomics 10: M111 010090

[20] Jaacks LM, Siegel KR, Gujral UP, Narayan KMV (2016) Type 2 diabetes: A 21st century epidemic. Best Practice & Research Clinical Endocrinology & Metabolism 30: 331-343

[21] Barrios C, Zierer J, Gudelj I, et al. (2016) Glycosylation Profile of IgG in Moderate Kidney Dysfunction. J Am Soc Nephrol 27: 933-941

[22] RDC T (2008) A language and environment for statistical computing. Available from <http://www.R-project.org>.

[23] Zaykin DV (2011) Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. J Evol Biol 24: 1836-1841

[24] Ercan A, Kohrt WM, Cui J, et al. (2017) Estrogens regulate glycosylation of IgG in women and men. JCI Insight 2

[25] Karsten CM, Pandey MK, Figge J, et al. (2012) Anti-inflammatory activity of IgG1 mediated by Fc galactosylation and association of FcgammaRIIB and dectin-1. Nat Med 18: 1401-1406

[26] Masuda K, Kubota T, Kaneko E, et al. (2007) Enhanced binding affinity for FcgammaRIIIa of fucose-negative antibody is sufficient to induce maximal antibody-dependent cellular cytotoxicity. Mol Immunol 44: 3122-3131

[27] Mihai S, Nimmerjahn F (2013) The role of Fc receptors and complement in autoimmunity. Autoimmun Rev 12: 657-660

[28] Nimmerjahn F, Ravetch JV (2008) Fcgamma receptors as regulators of immune responses. Nat Rev Immunol 8: 34-47

[29] Parekh RB, Roitt IM, Isenberg DA, Dwek RA, Ansell BM, Rademacher TW (1988) Galactosylation of IgG associated oligosaccharides: reduction in patients with adult and juvenile onset rheumatoid arthritis and relation to disease activity. Lancet 1: 966-969

[30] Parekh RB, Dwek RA, Sutton BJ, et al. (1985) Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. Nature 316: 452-457

[31] Vuckovic F, Kristic J, Gudelj I, et al. (2015) Association of systemic lupus erythematosus with decreased immunosuppressive potential of the IgG glycome. Arthritis Rheumatol 67: 2978-2989

[32] Trbojevic Akmacic I, Ventham NT, Theodoratou E, et al. (2015) Inflammatory bowel disease associates with proinflammatory potential of the immunoglobulin G glycome. Inflamm Bowel Dis 21: 1237-1247

[33] Vučković F, Theodoratou E, Thaçi K, et al. (2016) IgG Glycome in Colorectal Cancer. Clinical Cancer Research 22: 3078-3086

[34] Russell AC, Šimurina M, Garcia MT, et al. (2017) The N-glycosylation of immunoglobulin G as a novel biomarker of Parkinson's disease. Glycobiology 27: 501-510

[35] Dall’Olio F, Vanhooren V, Chen CC, Slagboom PE, Wuhrer M, Franceschi C (2013) N-glycomic biomarkers of biological aging and longevity: A link with inflammaging. Ageing Research Reviews 12: 685-698

[36] Ferrara C, Stuart F, Sondermann P, Brunker P, Umana P (2006) The carbohydrate at FcgammaRIIIa Asn-162. An element required for high affinity binding to non-fucosylated IgG glycoforms. J Biol Chem 281: 5032-5036

[37] Davies J, Jiang L, Pan LZ, LaBarre MJ, Anderson D, Reff M (2001) Expression of GnTIII in a recombinant anti-CD20 CHO production cell line: Expression of antibodies with altered glycoforms leads to an increase in ADCC through higher affinity for FC gamma RIII. Biotechnol Bioeng 74: 288-294

[38] Lauc G, Huffman JE, Pucic M, et al. (2013) Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. PLoS Genet 9: e1003225

[39] Testa R, Vanhooren V, Bonfigli AR, et al. (2015) N-glycomic changes in serum proteins in type 2 diabetes mellitus correlate with complications and with metabolic syndrome parameters. PLoS One 10: e0119983

[40] Itoh N, Sakaue S, Nakagawa H, et al. (2007) Analysis of N-glycan in serum glycoproteins from db/db mice and humans with type 2 diabetes. Am J Physiol Endocrinol Metab 293: E1069-1077

[41] Ercan A, Cui J, Chatterton DE, et al. (2010) Aberrant IgG galactosylation precedes disease onset, correlates with disease activity, and is prevalent in autoantibodies in rheumatoid arthritis. Arthritis Rheum 62: 2239-2248

[42] Menni C, Keser T, Mangino M, et al. (2013) Glycosylation of immunoglobulin g: role of genetic and epigenetic influences. PLoS One 8: e82558

[43] Lauc G, Essafi A, Huffman JE, et al. (2010) Genomics meets glycomics-the first GWAS study of human N-Glycome identifies HNF1alpha as a master regulator of plasma protein fucosylation. PLoS Genet 6: e1001256

[44] Kooner JS, Saleheen D, Sim X, et al. (2011) Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. Nat Genet 43: 984-989

[45] Hotamisligil GS (2010) Endoplasmic Reticulum Stress and the Inflammatory Basis of Metabolic Disease. Cell 140: 900-917

[46] Donath MY, Shoelson SE (2011) Type 2 diabetes as an inflammatory disease 11: 98-107

[47] Goldberg RB (2009) Cytokine and Cytokine-Like Inflammation Markers, Endothelial Dysfunction, and Imbalanced Coagulation in Development of Diabetes and Its Complications. The Journal of Clinical Endocrinology & Metabolism 94: 3171-3182

[48] Hertle E, Stehouwer CDA, van Greevenbroek MMJ (2014) The complement system in human cardiometabolic disease. Molecular Immunology 61: 135-148

[49] Ghosh P, Sahoo R, Vaidya A, Chorev M, Halperin JA (2015) Role of Complement and Complement Regulatory Proteins in the Complications of Diabetes. Endocrine Reviews 36: 272-288

[50] Holland M, Yagi H, Takahashi N, et al. (2006) Differential glycosylation of polyclonal IgG, IgG-Fc and IgG-Fab isolated from the sera of patients with ANCA-associated systemic vasculitis. Biochim Biophys Acta 1760: 669-677

**Table 1. Characteristics of the discovery and replication cohorts.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | DiaGene Study | | | CROATIA-Vis, CROATIA-Korcula, ORCADES | | |
| Cases (n=1815) | Controls (845) | p-value  cases vs controls | Cases (n=162) | Controls (n=3162) | p-value  cases vs controls |
| Age (yr.) | 65.2 (±10.5) | 65.7 (±6.8) | 0.198 | 66.6 (±10.6) | 54.5 (±14.9) | <0.001 |
| Sex (% male) | 53.7 | 40.2 | <0.001 | 45.7 | 38.5 | 0.082 |
| BMI (kg/m2) | 30.5 (±5.5) | 25.8 (±3.7) | <0.001 | 30.1 (±4.7) | 27.6 (±4.5) | <0.001 |
| HDL-cholesterol (mmol/L) | 1.17 (±0.32) | 1.46 (±0.36) | <0.001 | 1.38 (±0.37) | 1.52 (±0.39) | <0.001 |
| nonHDL-cholesterol (mmol/L) | 3.12 (±0.90) | 4.10 (±0.98) | <0.001 | 4.12 (±1.12) | 4.17 (±1.15) | 0.535 |
| Current smoking (%) | 17.8 | 10.9 |  | 14.8 | 16.3 |  |
| Former smoking (%) | 56.1 | 56.1 | <0.001\* | 31.5 | 30 | 0.852\* |
| Creatinine (umol/L) | 83.35 (±0.69) | 73.00 (±0.49) | <0.001 | 86.77 (±27.54) | 79.86 (±16.95) | 0.002 |

Unless stated otherwise, mean (±SD) are given. BMI, body mass index; HDL, high-density lipoprotein. \* p-value for trend across categories.

**Table 2: Correlations (continuous variables) and direction of effect in ANOVA (categorical variables) of clinical traits with IgG glycan peaks in the discovery population.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Glycan peak** | **Glycan structure** | **Age** | **Sex** | **BMI** | **Smoking** | **HDLc** | **NonHDLc** | **Creatinine** |
| **GP1** | FA1 | 0.26 | NS | NS | NS | NS | NS | 0.15 |
| **GP2** | A2 | 0.19 | ↑ | NS | NS | NS | NS | NS |
| **GP3** | A2B | 0.20 | NS | 0.12 | NS | -0.08 | NS | 0.10 |
| **GP4** | FA2 | 0.37 | ↑ | 0.15 | ↓ | -0.10 | -0.07 | 0.16 |
| **GP5** | M5 | 0.11 | NS | NS | NS | NS | NS | NS |
| **GP6** | FA2B | 0.37 | ↑ | 0.16 | ↑ | -0.12 | -0.11 | 0.16 |
| **GP7** | A2[3]G1 | NS | NS | -0.10 | NS | 0.08 | NS | NS |
| **GP8** | A2BG1 FA2[6]G1 | -0.23 | NS | -0.19 | NS | 0.17 | 0.13 | -0.04 |
| **GP9** | FA2[3]G1 | -0.15 | ↓ | -0.11 | ↓ | 0.10 | 0.10 | -0.16 |
| **GP10** | FA2[6]BG1 | NS | NS | NS | ↑ | NS | NS | NS |
| **GP11** | FA2[3]BG1 | 0.14 | NS | 0.08 | ↑ | NS | NS | 0.08 |
| **GP12** | A2G2 | -0.12 | NS | -0.08 | ↑ | NS | NS | NS |
| **GP13** | A2BG2 | -0.10 | NS | -0.07 | ↑ | NS | NS | NS |
| **GP14** | FA2G2 | -0.45 | ↓ | -0.11 | ↑ | NS | NS | -0.16 |
| **GP15** | FA2BG2 | -0.25 | ↓ | NS | ↑ | NS | NS | NS |
| **GP16** | FA2[6]G1S1 FA2[3]G1S1 | -0.09 | ↓ | NS | NS | NS | NS | NS |
| **GP17** | A2G2S1 | NS | NS | NS | NS | NS | NS | NS |
| **GP18** | A2BG2S1 FA2G2S1 | -0.41 | ↓ | -0.08 | NS | NS | NS | -0.14 |
| **GP19** | FA2BG2S1 | NS | NS | -0.15 | NS | NS | NS | NS |
| **GP20** | FA2FG2S1 | -0.10 | NS | NS | NS | NS | NS | NS |
| **GP21** | A2G2S2 | NS | NS | -0.08 | NS | NS | NS | NS |
| **GP22** | A2BG2S2 | NS | NS | NS | NS | NS | NS | NS |
| **GP23** | FA2G2S2 | -0.23 | ↓ | -0.09 | ↓ | NS | 0.07 | -0.12 |
| **GP24** | FA2BG2S2 | NS | NS | -0.11 | NS | 0.07 | NS | -0.08 |

βs are shown for continuous variables. Sex: ↑, higher in females; ↓, lower in females. Smoking: ↑, higher in current/former smokers; ↓, lower in current/former smokers. A p-value < 8.62E-4 was considered statistically significant. NS, not significant;

Structure abbreviations: F, α-1,6-linked core fucose; A, number of antenna's; B, bisecting GlcNac β1-4 linked to β1-3 mannose; M, number of mannose residues; Gx, number of β1-4 linked galactoses; [3]G1, galactose on the antenna of the α1-3 linked mannose; [6]G1, galactose on the antenna of the α1-6 linked mannose; Sx, number of sialic acids linked to galactose. Red background, positive association; blue background, negative association.

**Table 3. Statistically significant associations of IgG glycan traits for the discovery and replication cohorts and meta-analyses, adjusted for age and sex.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **Discovery** | | **Replication** | | **Meta-analysis** |
| **Glycan peaks** | | **β** | **p-value** | **β** | **p-value** | **p-value** |
| **GP3** | A2B | 4.81 | 5.90E-05 | 3.24 | 0.04 | 1.05E-05 |
| **GP4** | FA2 | 0.06 | 2.05E-14 | 0.01 | 0.73 | 8.97E-08 |
| **GP6** | FA2B | 0.39 | 1.21E-39 | 0.27 | 3.13E-06 | <2.22e-16 |
| **GP8** | A2BG1 | -0.35 | 1.88E-41 | -0.20 | 1.09E-05 | <2.22e-16 |
| **GP9** | FA2[3]G1 | -0.32 | 1.01E-23 | -0.23 | 1.40E-03 | <2.22e-16 |
| **GP10** | FA2[6]BG1 | 0.20 | 2.86E-08 | 0.33 | 8.02E-05 | 3.15E-11 |
| **GP11** | FA2[3]BG1 | 1.93 | 2.76E-10 | 1.23 | 8.63E-03 | 7.05E-10 |
| **GP14** | FA2G2 | -0.11 | 3.32E-12 | -0.01 | 0.88 | 1.95E-06 |
| **GP19** | FA2BG2S1 | -0.51 | 2.00E-06 | -0.23 | 0.18 | 3.31E-05 |
| **GP23** | FA2G2S2 | -0.63 | 3.08E-09 | -0.37 | 0.02 | 1.37E-08 |
|  |  |  |  |  |  |  |
| **Derived traits** | |  |  |  |  |  |
| **FBGS/(FBG+FBGS)** | Sialylation of fucosylated galactosylated structures with bisecting GlcNAc | -0.04 | 3.74E-08 | -0.05 | 1.00E-03 | 9.35E-10 |
| **FBGS/(FB+FBG+FBGS)** | Sialylation of fucosylated structures with bisecting GlcNAc | -0.09 | 1.54E-19 | -0.06 | 2.80E-04 | <2.22e-16 |
| **FG1S1/(FG1+FG1S1)** | S1 of fucosylated G1 structures | 0.21 | 1.26E-10 | 0.04 | 0.34 | 5.87E-07 |
| **FG2S1/(FG2+FG2S1+FG2S2)** | S1 of fucosylated G2 structures | 0.10 | 8.30E-09 | 0.03 | 0.44 | 9.87E-06 |
| **FBG2S1/(FBG2+FBG2S1+FBG2S2)** | S1 of fucosylated G2 structures with bisecting GlcNAc | -0.06 | 3.00E-06 | -0.05 | 0.02 | 1.52E-06 |
| **FS1/FS2** | Ratio of fucosylated S1 and S2 structures | 0.09 | 2.77E-04 | 0.14 | 1.00E-03 | 1.08E-06 |
| **FBS2/FS2** | Ratio of fucosylated S2 structures with and without bisecting GlcNAc | 0.86 | 5.11E-10 | 0.62 | 3.10E-03 | 2.16E-10 |
| **FBS2/(FS2+FBS2)** | B in all fucosylated S2 structures | 4.71 | 2.24E-10 | 4.17 | 5.30E-03 | 2.83E-10 |
| **G0n** | G0 structures in total neutral glycans | 0.07 | 1.56E-24 | 0.02 | 0.13 | 2.22E-15 |
| **G1n** | G1 structures in total neutral glycans | -0.19 | 3.55E-39 | -0.08 | 5.80E-04 | <2.22e-16 |
| **G2n** | G2 structures in total neutral glycans | -0.06 | 9.67E-10 | 0.01 | 0.79 | 1.06E-04 |
| **Fn** | F in neutral glycans (without bisecting GlcNAc) | -0.12 | 4.55E-20 | -0.11 | 8.10E-06 | <2.22e-16 |
| **FG0n/G0n** | F in neutral G0 structures | -0.06 | 3.18E-09 | -0.08 | 1.86E-05 | 9.38E-13 |
| **FG1n/G1n** | F in neutral G1 structures | -0.13 | 1.93E-24 | -0.13 | 2.03E-07 | <2.22e-16 |
| **FG2n/G2n** | F in neutral G2 structures | -0.10 | 1.19E-12 | -0.02 | 0.21 | 1.39E-08 |
| **FBn** | F in neutral structures with bisecting GlcNAc | 0.16 | 1.72E-26 | 0.18 | 8.74E-09 | <2.22e-16 |
| **FBG0n/G0n** | F in neutral G0 structures with bisecting GlcNAc | 0.09 | 1.10E-14 | 0.13 | 3.99E-07 | <2.22e-16 |
| **FBG1n/G1n** | F in neutral G1 structures with bisecting GlcNAc | 0.14 | 1.34E-26 | 0.16 | 1.78E-09 | <2.22e-16 |
| **FBG2n/G2n** | F in neutral G2 structures with bisecting GlcNAc | 0.22 | 8.01E-21 | 0.06 | 0.06 | 2.09E-14 |
| **FBn/Fn** | Ratio of fucosylated neutral glycans with and without bisecting GlcNAc | 0.10 | 2.73E-25 | 0.12 | 2.14E-06 | <2.22e-16 |
| **FBn/Fn total** | Ratio of bisecting GlcNAc in all fucosylated structures within neutral glycans | 0.15 | 1.63E-25 | 0.16 | 2.45E-08 | <2.22e-16 |

A p-value <8.62E-4 was considered statistically significant. Structure abbreviations: F, core fucose; A, number of antenna's; B, bisecting GlcNAc; M, number of mannose residues; Gx, number of galactoses; [3]G1, galactose on the α1-3 antenna; [6]G1, galactose on the α1-6 antenna; Sx, number of sialic acids linked to galactose; n, neutral glycans. Red: positive association with type 2 diabetes; blue: negative association.

**Table 4. Associations of IgG glycan traits for the discovery and replication cohorts and meta-analyses, adjusted for age, sex, BMI, HDL-cholesterol, non-HDL cholesterol, smoking and serum creatinine.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **Discovery cohort** | | **Replication cohort** | | **Meta-analysis** |
| **Glycan peaks** | | **β** | **p-value** | **β** | **p-value** | **p-value** |
| **GP6** | FA2B | 0.23 | 2.45E-08 | 0.24 | 3.89E-05 | 1.84E-11 |
| **GP8** | A2BG1 | -0.22 | 1.59E-09 | -0.20 | 2.10E-05 | 1.10E-12 |
| **GP9** | FA2[3]G1 | -0.19 | 4.30E-05 | -0.22 | 3.15E-03 | 1.08E-06 |
| **GP10** | FA2[6]BG1 | 0.14 | 6.56E-03 | 0.30 | 4.93E-04 | 9.96E-06 |
| **GP11** | FA2[3]BG1 | 1.28 | 3.10E-03 | 1.10 | 0.03 | 3.28E-04 |
|  |  |  |  |  |  |  |
| **Derived traits** | |  |  |  |  |  |
| **FBGS/(FB+FBG+FBGS)** | Sialylation of fucosylated structures with bisecting GlcNAc | -0.03 | 0.06 | -0.05 | 1.89E-03 | 3.28E-04 |
| **FBS2/FS2** | Ratio of fucosylated S2 structures with and without bisecting GlcNAc | 0.68 | 7.42E-04 | 0.59 | 5.89E-03 | 1.96E-05 |
| **FBS2/(FS2+FBS2)** | Bisecting GlcNAc in all fucosylated S2 structures | 3.72 | 6.27E-04 | 4.02 | 8.62E-03 | 2.67E-05 |
| **G1n** | G1 structures in total neutral glycans | -0.10 | 1.00E-06 | -0.07 | 2.70E-03 | 7.40E-08 |
| **Fn** | F in neutral glycans (without bisecting GlcNAc) | -0.08 | 6.00E-06 | -0.02 | 2.37E-05 | 8.90E-10 |
| **FG0n/G0n** | F in neutral G0 structures | -0.05 | 4.06E-04 | -0.08 | 2.76E-05 | 4.16E-08 |
| **FG1n/G1n** | F in neutral G1 structures | -0.08 | 2.00E-06 | -0.12 | 1.92E-06 | 2.36E-11 |
| **FBn** | F in neutral structures with bisecting GlcNAc | 0.11 | 2.09E-07 | 0.17 | 9.07E-08 | 1.13E-13 |
| **FBG0n/G0n** | F in neutral G0 structures with bisecting GlcNAc | 0.08 | 1.00E-05 | 0.13 | 5.62E-07 | 2.65E-11 |
| **FBG1n/G1n** | F in neutral G1 structures with bisecting GlcNAc | 0.09 | 9.91E-07 | 0.16 | 4.08E-08 | 2.04E-13 |
| **FBG2n/G2n** | F in neutral G2 structures with bisecting GlcNAc | 0.13 | 8.10E-05 | 0.06 | 0.08 | 1.24E-04 |
| **FBn/Fn** | Ratio of fucosylated neutral glycans with and without bisecting GlcNAc | 0.07 | 3.79E-07 | 0.11 | 1.54E-05 | 5.25E-11 |
| **FBn/Fn total** | Ratio of bisecting GlcNAc in all fucosylated structures within neutral glycans | 0.10 | 3.42E-07 | 0.16 | 1.80E-07 | 3.63E-13 |

A p-value <8.62E-4 was considered statistically significant. Structure abbreviations: F, core fucose; A, number of antenna's; B, bisecting GlcNAc; M, number of mannose residues; Gx, number of galactoses; [3]G1, galactose on the α1-3 antenna; [6]G1, galactose on the α1-6 antenna; Sx, number of sialic acids linked to galactose; n, neutral glycans. Red: positive association with type 2 diabetes; blue: negative association.

**Table 5. Ten-fold cross validated AUCs standardized.**

|  |  |  |
| --- | --- | --- |
| Model | AUC without 24 IgG glycans | AUC with 24 IgG glycans |
| DiaGene study  Glycans only  Age, sex  Age, sex, BMI  Age, sex, BMI, HDL-c, non-HDL-c, smoking  Age, sex, BMI, HDL-c, non-HDL-c, creatinine, smoking  CROATIA-Vis, CROATIA-Korcula, ORCADES populationa  Glycans only  Age, sex  Age, sex, BMI  Age, sex, BMI, HDL-c, non-HDL-c, smoking  Age, sex, BMI, HDL-c, non-HDL-c, creatinine, smoking | 0.54  0.78  0.89  0.89  0.77  0.79  0.81  0.81 | 0.73  0.73  0.82  0.89  0.90  0.75  0.80  0.82  0.82  0.82 |

All analyses in CROATIA-Vis, CROATIA-Korcula and ORCADES were adjusted for cohort ID. BMI, body mass index; HDL-c, high-density lipoprotein-cholesterol; non-HDL-c, non-high-density lipoprotein cholesterol.**Figure 1. The 24 IgG glycan peaks as measured with ultra-performance liquid chromatography.** Shown are the structure of the main IgG glycan component(s) per peak and their structure abbreviation. Structure abbreviations: F, α-1,6-linked core fucose; A, number of antenna's attached to the core sequence (existing of two N-Acetylglucosamine (GlcNAc) and three mannose residues); B, bisecting GlcNac β1-4 linked to β1-3 mannose; M, number of mannose residues; G, number of β1-4 linked galactoses; [3]G1, galactose on the antenna of the α1-3 linked mannose; [6]G1, galactose on the antenna of the α1-6 linked mannose; S, sialic acid linked to galactose. Structural schemes are defined as follows: blue square, GlcNac; green circle, mannose; red triangle, core fucose; yellow circle, galactose; purple rhomb, sialic acid.

**Figure 2. UPLC chromatogram of one individual with and one individual without type 2 diabetes.** The 24 IgG glycan peaks are numbered and peaks associated with type 2 diabetes in the full model are highlighted in red. Black line, healthy control; blue line, individual with type 2 diabetes.

**Figure 3. Distribution of IgG glycan peaks associated with type 2 diabetes after adjustment for the full model.** Upper half: distribution of glycan peaks in the discovery cohort. Bottom half: distribution of glycan peaks in the replication cohort. Shown are the relative percentages of each peak for controls (grey) and cases (blue). GP, glycan peak.

**Figure 4. ROC curves for type 2 diabetes.** The AUC is shown for 1) IgG glycans alone (green); 2) age and sex (red); 3) age, sex, and IgG glycans (blue) in A for the discovery cohort and in B for the replication cohort. AUCs for 4) the full model (red); and 5) the full model with IgG glycans (blue) are shown in C for the discovery cohort and in D for the replication cohort. AUC, area under the curve.

**Supplementary table S1. Description of direct glycan peaks and derived traits measured in this study.** Structure abbreviations are as follows: F, α-1,6-linked core fucose; A, number of antenna's; B, bisecting GlcNac β1-4 linked to β1-3 mannose; M, number of mannose residues; Gx, number of β1-4 linked galactoses; [3]G1, galactose on the antenna of the α1-3 linked mannose; [6]G1, galactose on the antenna of the α1-6 linked mannose; Sx, number of sialic acids linked to galactose.

**Supplementary table S2. Median and interquartile ranges of glycan peaks for cases and controls in the discovery and replication cohort.** GP, glycan peak; IQR, interquartile range. \* In the replication cohort, GP3 was only measured in the ORCADES study but not in the CROATIA-Vis or -Korcula studies. These values thus only represent GP3 in the ORCADES study.

**Supplementary table S3. Associations of IgG glycans and derived traits adjusted for age and sex.** Glycan peaks and derived traits that were significantly associated with type 2 diabetes in the meta-analysis are in bold font. Structure abbreviations are as follows: F, α-1,6-linked core fucose; A, number of antenna's; B, bisecting GlcNac β1-4 linked to β1-3 mannose; M, number of mannose residues; Gx, number of β1-4 linked galactoses; [3]G1, galactose on the antenna of the α1-3 linked mannose; [6]G1, galactose on the antenna of the α1-6 linked mannose; Sx, number of sialic acids linked to galactose. Red: positive association between glycan and clinical trait; blue: negative association.

**Supplementary Table S4. Associations of IgG glycans and derived traits adjusted for age, sex and BMI.** Glycan peaks and derived traits that were significantly associated with type 2 diabetes in the meta-analysis are in bold font. Structure abbreviations are as follows: F, α-1,6-linked core fucose; A, number of antenna's; B, bisecting GlcNac β1-4 linked to β1-3 mannose; M, number of mannose residues; Gx, number of β1-4 linked galactoses; [3]G1, galactose on the antenna of the α1-3 linked mannose; [6]G1, galactose on the antenna of the α1-6 linked mannose; Sx, number of sialic acids linked to galactose. Red: positive association between glycan and clinical trait; blue: negative association.

**Supplementary Table S5. Associations of IgG glycans and derived traits adjusted for age, sex, BMI, smoking, HDL-cholesterol, non-HDL-cholesterol, smoking and serum creatinine.** Glycan peaks and derived traits that were significantly associated with type 2 diabetes in the meta-analysis are in bold font (p-value <8.62E-4). Structure abbreviations are as follows: F, α-1,6-linked core fucose; A, number of antenna's; B, bisecting GlcNac β1-4 linked to β1-3 mannose; M, number of mannose residues; Gx, number of β1-4 linked galactoses; [3]G1, galactose on the antenna of the α1-3 linked mannose; [6]G1, galactose on the antenna of the α1-6 linked mannose; Sx, number of sialic acids linked to galactose. Red: positive association between glycan and clinical trait; blue: negative association. S.E., standard error.

**Supplementary figure S1.** **Correlation between measurements of replicated samples in the Discovery cohort (69 samples out of 2688) per glycan peak.**

**Supplementary figure S2. Variation in measurements of standard samples in the Discovery cohort (n=153, ca. 5 samples per plate) per glycan peak.**