Investigation of the causal relationships between human IgG N-glycosylation and twelve common diseases associated with changes in the IgG N-glycome

Investigation of the causal relationships between human IgG N-glycosylation and twelve common diseases associated with changes in the IgG N-glycome

Olga O. Zaytseva[1], Sodbo Zh. Sharapov[2], Marcus Perola[3], Tonu Esko[4], Arianna Landini[5], Caroline Hayward[6], James F. Wilson[5,6], Gordan Lauc[1,7], Yurii S. Aulchenko[2], Lucija Klarić[6], Yakov A. Tsepilov[8,9]

[1] Genos Glycoscience Research Laboratory, Zagreb, 10000, Croatia

[2] Laboratory of Glycogenomics, Institute of Cytology and Genetics, Novosibirsk, 630090, Russia

[3] Genomics and Biomarkers Unit, Department of Health, National Institute for Health and Welfare (THL), Helsinki, FI-00271, Finland

[4] Estonian Genome Center, University of Tartu, Tartu, 51010, Estonia

[5] Centre for Global Health Research, Usher Institute, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG, Scotland

[6] MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, EH4 2XU, United Kingdom

[7] Faculty of Pharmacy and Biochemistry, University of Zagreb, Zagreb, 10000, Croatia

[8] Laboratory of Theoretical and Applied Functional Genomics, Novosibirsk State University, Novosibirsk, 630090, Russia

[9] Laboratory of Recombination and Segregation Analysis, Institute of Cytology and Genetics, Novosibirsk, 630090, Russia

Correspondence to:
Olga Zaytseva (ozaitseva@genos.hr)

Abstract

Changes in the N-glycosylation of immunoglobulin G (IgG) are often observed in pathological states, such as autoimmune, inflammatory, neurodegenerative, cardiovascular diseases and some types of cancer. However, in most cases it is not clear if the disease onset causes these changes, or if the changes in IgG N-glycosylation are among the risk factors for the diseases. The aim of this study was to investigate the casual relationships between IgG N-glycosylation traits and 12 diseases, in which the alterations of IgG N-glycome were previously reported, using Two Sample Mendelian Randomization (MR) approach. We have performed Two Sample MR using publicly available summary statistics of genome-wide association studies of IgG N-glycosylation and disease risks. Our results indicate positive causal effect of systemic lupus erythematosus (SLE) on the abundance of N-glycans with bisecting *N*-acetylglucosamine in the total IgG N-glycome. Therefore, we suggest regarding this IgG glycosylation trait as a biomarker of SLE. We also emphasize the need for more powerful GWAS studies of IgG N-glycosylation to further elucidate the causal effect of IgG N-glycome on the diseases.

Keywords: N-glycosylation, immunoglobulin G, two sample Mendelian randomization, systemic lupus erythematous, disease risk

Introduction

In recent years, significant attention has been brought to N-glycans as biomarkers of various pathological states (1–5), as well as of aging (6,7). N-glycosylation, posttranslational attachment of oligosaccharides to asparagine residues of the protein backbone, is known to influence physical, chemical and biological properties of the modified protein(8,9). Immunoglobulin G (IgG) is among the most abundant N-glycosylated proteins in the healthy human blood plasma(10) and is one of the proteins with the best-studied spectrum of possible N-glycan modifications (N-glycome). The majority of IgG N-glycans consist of a conserved core with two *N*-acetylglucosamine (GlcNAc) antennae, and may contain core-fucose, bisecting GlcNAc, and antennary galactosylation and/or sialylation (11) (**Fig. 1**). IgG N-glycosylation is altered in ageing (7); ovarian(12), prostate(13), breast (14), colorectal (15) and lung (16) cancers; inflammatory and autoimmune diseases, such as rheumatoid arthritis (RA)(4,17), systemic lupus erythematous (SLE)(18), inflammatory bowel disease (IBD)(19), neurodegenerative diseases such as Parkinson's disease (PD)(20) and Alzheimer's disease (AD)(21); type 1(22) and type 2(23) diabetes; hypertension(24) and others (reviewed in (25)). The putative mechanisms through which IgG N-glycosylation acts in diseases include differential engagement of Fcγ-receptors(26–28) and complement system (29) depending on the structure of N-glycans attached to the effector domain of IgG. In this way IgG N-glycome affects the path of immune response and contributes to pathological processes, as happens, for instance, in RA(30).

N-glycan synthesis in the cell is not template-driven like that of nucleic acids and proteins. The biochemical pathways of N-glycan synthesis are complicated and include many glycosyltransferase and glycosidase enzymes, whose expression and activity are controlled by diverse factors, including Golgi pH and ionic composition(31), properties of Golgi membranes(32), competition between enzymes, abundance of activated sugars, which are building blocks for the glycan chains(33). Dozens of genetic variants have been discovered that influence IgG N-glycome composition (34–37). A number of loci associated with IgG N-glycome colocalize with loci implicated in autoimmune and inflammatory diseases (36,37). Recently, N-glycome of IgG has been regarded as a potential biomarker, reflecting changes in health and disease progression(25). However, there is little knowledge about causal relationships between IgG N-glycosylation and disease risk. In most cases, it is not clear whether the abnormal glycosylation profile of IgG is involved in disease onset or if it is a consequence of the pathologic processes in the organism.

The gold standard epidemiological design used for analysis of causal relationships is a randomized controlled trial, which in case of IgG N-glycosylation and disease risk in humans would be very challenging. A Mendelian randomization (MR) approach uses genetic variants that influence traits in question as instruments for making causal inferences (38). In order to test if one trait (the exposure) has a causal effect on the other (the outcome), SNPs that are known to affect the exposure are selected as instrumental variables (IVs), and their effects both on exposure and outcome are used to estimate the causal effect(39,40). Effects of the genetic variants can be obtained from GWAS on the respective traits. To be able to perform MR, one does not need to have genetic associations for both traits analysed in the same cohort of people, the presence of selected genetic variants, or reliable LD proxies, in both cohorts is enough. This approach is usually referred to as Two-Sample MR (39).

In this study we describe the first large-scale investigation of causal relationships between IgG N-glycosylation traits and risk of 12 common diseases, whose aberrant glycosylation profile is well characterised. The 12 diseases were selected on the basis that IgG N-glycosylation was reported to be altered in these conditions (25) and, in some cases, shared genetic variants existed(36,37), influencing both IgG N-glycosylation and the disease. We used summary statistics from the largest previously published GWAS of IgG N-glycosylation traits (36,37) and 12 autoimmune, inflammatory, neurodegenerative, cardiovascular and cancer diseases to estimate the causal effect of IgG N-glycans on diseases and *vice versa,* using the Two-Sample MR approach (39). The diseases studied included IBD, Crohn's disease (CD) and ulcerative colitis (UC)(41), type 2 diabetes (T2D)(42), coronary artery disease (CAD)(43), hypertension (http://www.nealelab.is), asthma (http://www.nealelab.is), RA (44), SLE(45), AD (46), PD (47) and lung cancer(48). We found support for the positive causal effect of SLE on the abundance of bisected N-glycans attached to IgG. Moreover, we replicated the observed effect using a multivariate MR approach.

Results

The general scheme of the analysis is depicted in **Fig. 2**. More details about the IVs selection and sensitivity analyses could be found in the Methods and Materials section.

Genetic basis of IgG N-glycosylation traits that we considered for this study was previously investigated in (35–37). These 86 traits describe either relative abundances of individual N-glycans found attached to IgG (glycan traits IGP1-23) or abundances of various N-glycans sharing similar

structural characteristics and their ratios (IGP24-77 and nine compound glycosylation traits). The list of analysed traits with descriptions is available in **Supplementary Table S1 A, B**.

The twelve various autoimmune, inflammatory, neurodegenerative, cardiovascular and cancer diseases (**Table 3**) that are characterized by the alteration of the IgG N-glycome (25) were selected for the study if summary statistics for the GWAS of the trait was publicly available via either the TwoSampleMR" R package(39) or the GWAS-MAP data base (49,50).

Causal effect of IgG N-glycosylation traits on the diseases

To test for the causal effect of IgG N-glycosylation traits on the selected 12 diseases we used the summary statistics from the largest available GWAMA for each of the 86 N-glycan traits, *i.e.*, 9K GWAMA. We applied rigorous SNP selection to ensure the robust association of the SNP with the exposure (see Methods and Materials for the details of the IV selection). Each SNP selected as an IV had F-statistic > 10 and was considered as a strong IV. Seventy-two (72) IgG N-glycosylation traits provided at least two valid IVs robustly associated with the trait in a corresponding 9K GWAMA. Maximum number of IVs associated with an IgG N-glycosylation trait was 9 for the following IgG N-glycosylation traits: IGP14, IGP40, IGP49, IGP67, IGP68, IGP69 and IGP75. Median and mean numbers of IVs per IgG N-glycosylation trait GWAS in this set of 72 traits were 5.43 and 6, respectively (**Supplementary Table S2**). The selected IVs on average explained 5.3% of variance in IgG N-glycosylation traits (minimal value of 0.67% variance in IGP25 explained, maximal value of 20.06% variance in IGP29 explained), with an average F-statistics of an IV set associated with an individual IgG N-glycan trait of 476.02, which indicated sufficient average IV set strength (**Supplementary Table S2**). Each of the 72 IgG N-glycosylation traits was tested for causal effect on each of the 12 diseases. The results of the corresponding Two-Sample MR analyses can be found in **Supplementary Table S3.**

None of the analysed IgG N-glycosylation trait-disease pairs produced any causal effect estimates that would be statistically significant at our designated threshold of $p \leq 1.98 \times 10^{-4}$ (multiple correction for 21 principal components of IgG N-glycosylation traits *12 diseases). Nevertheless, we observed one suggestive causal signal of the effect of the IgG N-glycan trait IGP76 (ratio of fucosylated digalactosylated non-bisected structures and all digalactosylated structures with bisecting GlcNAc) on IBD ($p = 3.66 \times 10^{-4}$). Corresponding IVs are listed in **Supplementary Table S4**. The effect was estimated to be 0.18 log odds ratio units of IBD per standard deviation units of IGP76 with inverse variance weighted (IVW) regression (**Table 1**, **Supplementary Fig. S1 A, B, E**). Leave-one-out sensitivity analysis did not indicate the overall effect being driven by a

single SNP (**Supplementary Fig. S1 C)**. The funnel plot that is used to visualize the directional bias of weaker IVs appears asymmetrical, though it might be due to a small number (six) of IVs (**Supplementary Fig. S1 D)**. No heterogeneity was observed between IVs either with IVW (Q = 3.40, p =0.64) or MR Egger regression (Q = 2.62; p = 0.62) and the MR Egger intercept was not significantly different from 0 (intercept estimate 0.022, se = 0.025, p = 0.43, **Supplementary Fig S1 E**). Estimates with other MR methods, namely, MR Egger, Weighted median, Simple mode and Weighted mode were directionally consistent, however none of the estimates apart from IWV were statistically significant (**Table 1, Supplementary Figs S1 B**). As additional sensitivity analysis we performed Two-Sample MR of IGP76 on IBD having excluded two outlier IVs identified with the MR-PRESSO(51) R package from the analysis (see **Supplementary Table S4**) and observed an estimate of causal effect of the same directionality and similar magnitude with IVW, 0.137 (p = 0.023), log odds units of IBD risk per standard deviation unit of IGP76 (**Table 1**, **Supplementary Fig. S2 A, B, E).** Leave-one-out analysis did not indicate the overall effect being driven by a single SNP (**Supplementary Fig. S2 C)**, funnel plot nevertheless showing asymmetric distribution of SNPs (**Supplementary Fig. S2 D)**. Other MR methods produced estimates of same directionality and similar magnitudes, with no evidence for directional horizontal pleiotropy (no significant heterogeneity among IVs in both IVW (Q = 1.40, p = 0.71) and MR Egger (Q = 1.22, p = 0.54), intercept not statistically significantly different from 0 in MR Egger regression, see **Supplementary Figs S2 E**). However, none of these causal estimates were reaching statistical significance at $p \leq 1.98 \times 10^{-4}$ (p ranging from 0.023 with IVW regression to 0.179 with MR Egger regression). We were unable to conduct replication of this causal association since no GWAS of IBD performed in an independent cohort was publicly available.

Causal effect of the diseases on the IgG N-glycosylation traits

In the discovery round analysing the causal effect of complex diseases on IgG glycosylation for each of the 12 diseases included we selected independent IVs associated with the disease at the genome-wide level ($p \leq 5 \times 10^{-8}$, clumped within a window of 500 Kb) with MAF ≥ 0.05 in the summary statistics. Information on the strength of the IVs is presented in **Supplementary Table S5**. The average number of IVs per disease was 51, with maximum number of IVs (160) for hypertension and minimum of 6 for PD. The average strength of the IVs associated with diseases expressed in terms of F-statistics was 3811 (maximum of 8795 for hypertension and minimum of 285 for PD). On average the selected IVs explained 7.6% of variance in disease. As outcomes we used summary

statistics of each of the 86 IgG N-glycosylation traits in the 8K cohort. The estimates of causal effect for the disease - glycosylation trait pairs can be found in **Supplementary Table S6**.

We obtained a significant causal association of SLE risk with bisection of IgG N-glycans, IgG_B ($p = 5.12x10^{-7}$, **Table 2, Supplementary Fig. S3 A, B, E**) at the designated significance threshold of $p \leq 1.98x10^{-4}$ with IVW regression using 36 independent genetic variants explaining associated with SLE as IVs. The effect was estimated to be 0.123 standard deviation units of IgG_B per log odds ratio unit of SLE. Estimates with other MR methods were directionally consistent and similar in magnitude, however not statistically significant at the designated threshold (**Table 2**, **Supplementary Fig. S3 A, B, E**). The funnel plot was only slightly asymmetric (**Supplementary Fig. S3 C**), and the leave-one-out analysis showed that no single IV was driving the effect observed (**Supplementary Fig. S3 D).** MR Egger intercept was not statistically significantly different from 0 (intercept = -0.011, p = 0.197) and both MR Egger (Q = 44.56, p = 0.11) and IVW (Q = 46.83, p = 0.09) heterogeneity tests suggested there are no outliers (**Supplementary Fig. S3 E)**.

For the sensitivity analysis of the causal estimate of SLE effect on IgG bisection we have curated the list of instruments. This time we selected 41 independent SNPs that were found to be significantly associated with SLE in (45), or the available proxy SNPs (see Materials and Methods, **Supplementary Table S7**). None of the selected IVs was previously reported to be associated with the IgG N-glycome. The MR-PRESSO software then identified 5 out of 41 instruments as outliers (**Supplementary Table S7**). Two-Sample MR was then performed on the refined set of 36 IVs. Summary statistics source for the exposure was the same as those used for discovery; as outcome summary statistics we used the 8K IgG_B GWAS GWAMA. The casual estimate of SLE on bisection of IgG glycans was found to be 0.124 standard deviation units of IgG_B per log odds ratio units of SLE at $p = 1.24x10^{-5}$ (**Table 2, Supplementary Fig. S4 A, B, E**) which is close to the magnitude of the effect estimated in the discovery MR. Estimates with other MR methods were directionally consistent and similar in magnitude (**Table 2, Supplementary Fig. S4 A, B, E).** The funnel plot looked more symmetric (**Supplementary Fig. S4 C)** and leave-one analysis suggested no single IV driving the estimated causal effect (**Supplementary Fig. S4 D).** No significant heterogeneity was detected both with MR Egger (Q = 40.20, p = 0.22) and IVW (Q = 40.46, p = 0.24) regressions, and MR Egger intercept was not statistically different from 0 (intercept = -0.003, p = 0.643) **(see Supplementary Fig. S4 E).** For the univariate replication as the exposure we used summary statistics for the same set of 36 SNPs as in Sensitivity 2 analysis in the SLE GWAS (45) and summary statistics for bisection of IgG glycans from the 3K GWAMA. However, in this MR analysis the causal estimate turned out to be not significant (**Table 2**, Supplementary **Fig. S5 A, B, E**) with any of the

used regression methods (with IVW: causal effect of -0.016 standard deviation units of IgG_B per log odds ratio units of SLE at p = 0.72), in the absence of heterogeneity with MR Egger (Q = 24.67, p = 0.88), Maximum Likelihood (Q = 24.72, p = 0.90) and IVW (Q = 24.72, p= 0.90) methods (**Supplementary Fig. S5 C),** leave-one-out sensitivity analysis did not indicate the overall effect being driven by a single SNP **(Supplementary Fig. S5 D)** and no significant horizontal pleiotropy detected with MR Egger estimates of the intercept (intercept = -0.002, p = 0.819, **Supplementary Fig. S5 E**).

Multivariate MR Analysis

There are two main explanations why we have not replicated the signal related to the causal effect of SLE on IgG_B in the univariate replication. The first one is the power issue – our outcome replication set was 2.5 times smaller than the discovery. The second one is the property of the outcome (IgG_B). As there are no available GWAS for IgG_B trait itself, we have obtained the GWAS summary statistics for IgG_B as the weighted linear combination of the GWAS of the original 23 IGPs with directly calculated GWAS. Such linear summary level combination increases the error of the effect estimates and hence decreases the power of replication.

As an additional attempt to replicate the promising causal signal of SLE on IgG bisection we adapted the multivariate ANOVA (MANOVA) approach to MR(52). The MV approach to Two Sample MR takes advantage from the fact that the 86 IgG N-glycosylation traits analysed in this study are correlated. Multivariate analysis of a group of correlated phenotypes when the correlation structure is considered can increase the statistical power of the study(35,52). Relative abundance of bisected glycans in the IgG N-glycome, from this point of view can be approximated as a combination of the individual N-glycans that are correlated with IgG_B. From the set of the 23 IgG N-glycosylation traits (IGP1-23) that represent directly measured relative abundances of specific N-glycan structures in the IgG N-glycome we selected those correlated with IgG_B. Then we estimated the causal effect of SLE on each IGP correlated with IgG_B. Then we applied the MANOVA approach to make an estimate of the causal effect of SLE on the IgG_B-correlated N-glycans. The results of these Multivariate (MV) MR are presented in **Table 2**.

The causal effect of SLE on the group of N-glycans correlated with IgG bisection in the Discovery MV analysis was significant when the 8K GWAMA summary statistics was used as outcomes ($p=-9.43 \times 10^{-4}$), which proved the MV MR approach applicable in this particular case. The signal was again statistically significant at the 0.05 significance level in the MV MR Replication Stage 1, when

summary statistics for the outcomes was taken from the 3K GWAMA with $p = 5.98 \times 10^{-4}$. At the MV MR Replication Stage 2 the coefficients for the MANOVA analysis were taken from the results of the MV MR Discovery, and the 3K GWAMA was used as a source of summary statistics for the outcomes. The signal in the Stage 2 of MV MR replication again was statistically significant ($p = 0.003$) at 0.05 significance level and directionally consistent with the effect estimated in the Discovery Stage of MV MR. Therefore, we consider the causal effect of SLE risk on bisection of the IgG N-glycans replicated in the MV MR analysis.

Discussion

Uncovering causal relationships between IgG N-glycosylation and disease risk can be beneficial for biomarker discovery and provide new treatment options. It can also enrich our understanding of the mechanisms of diseases and shed new light on the role of the IgG N-glycome in immunity. In the current study we aimed to investigate the causal relationships between the risk of 12 autoimmune, cardiovascular, neurodegenerative, or/and inflammatory diseases, as well as lung cancer, and IgG N-glycosylation features using the Two-Sample MR approach. To our knowledge, our study is the first one to apply the MR method to evaluate the causal relationships between IgG N-glycome and disease risk using the most extensive list of genetic associations with IgG N-glycosylation, using 29 genetics variants as IVs (36,37). A previous study reported no significant association of genetic variants involved in IgG glycosylation with RA(53), however it used only 16 SNPs of out of 17 genetic variants identified in an early work (34) published in 2013.

When we performed MR analysis regarding IgG N-glycosylation traits as exposures, we were unable to detect statistically significant causal effects on investigated diseases. The only suggestive signal that we observed was related to the effect of increased IGP76 trait leading to higher risk of IBD. IGP76 is a trait describing the ratio of fucosylated digalactosylated non-bisected structures and all digalactosylated structures with bisecting GlcNAc. IGP76 is inversely proportional to the overall abundance of bisected structures in the IgG N-glycome (phenotypic correlation -0.73, see **Supplementary Fig. S6**). Bisection of IgG N-glycans is known to increase in IBD patients as compared with healthy controls, while galactosylation decreases (54), which contradicts the direction observed in our study. Unfortunately, we could not replicate this finding in the current work, since no GWAS of IBD performed in an independent cohort was publicly available at the moment when the study was carried out. Further studies are required to clarify the relationships between IGP76 and IBD.

Analysing the causal effect of the chosen diseases on the IgG N-glycome composition we found evidence of SLE leading to increase in the percentage of bisected structures in the IgG N-glycome when summary statistics from GWAS of IgG bisection in the discovery cohort was used. However, we were unable to replicate this signal, neither in direction, nor in magnitude, using 2.5 times smaller replication cohort summary statistics as outcome in the Two Sample MR. It is important to remember, that directly conducted GWAS bisection of IgG N-glycans is not currently available, and neither are GWAS for such derived traits as proportion of glycans with core-fucose, with antennary galactosylation, sialylation, *etc*. At the same time, such traits describe the very basic characteristics of the IgG N-glycome and are commonly reported in the functional and case-control studies of IgG glycosylation (for example, see (25)). To add these traits to our study we have constructed GWAS summary statistics for such general traits, including IgG_B, by combining the effects observed for individual glycans that are summed up to form the trait (for example, relative abundances of every bisected N-glycan species are summed to provide the proportion of bisected glycans in the IgG N-glycome). In brief, we were combining the effects of a taken genetic variant on the abundance of every bisected glycan species considering the corresponding variances of the measured glycan abundances, phenotypic correlations between these individual glycan abundances and the standard errors associated with the effects. This allowed us to apply MR methods to these general traits as well, but naturally the resulting summary statistics suffered from power loss and the inability to replicate the causal effect of SLE on IgG_B in the 3K cohort might have been due to it. To increase the power of the analysis we have therefore also applied the multivariate approach to the MR analysis of SLE effect on bisection of IgG N-glycans. In essence, the point of the multivariate MR was to test whether the effects of SLE on each of the individual glycan species that are correlated with IgG_B are consistent with the effect of SLE on IgG_B, when we take into account the correlation structure in this group of traits. We obtained statistically significant estimates of the causal effect both in multivariate discovery, as well as in Stages 1 and 2 of MV MR replication. Thus, we considered the causal effect of SLE on IgG_B replicated in our study, although this result still needs to be interpreted with caution. Since we have performed tests for all the pairs of IgG N-glycosylation traits and diseases in both directions and no significant causal effect for IgG N-glycan bisection on SLE was observed, we assume that it serves as additional confirmation of the absence of reverse causality.

When we consider what is known from case-control studies about the alterations of the IgG N-glycome in SLE, our results seem to fit into the biological context quite straightforwardly. Bisection of IgG is known to be elevated in SLE (18), as well as in other autoimmune and inflammatory

diseases, such as RA, IBD (25) and is associated with proinflammatory state (55). Nevertheless, the functional role of bisected IgG glycans in these processes is unclear (56). We can assume that increased bisection in SLE might be a consequence of lower fucosylation of IgG N-glycans (18), since bisection and core-fucosylation of N-glycans are opposing negatively correlated processes (57). However, we did not observe any statistically significant causal relationships between SLE and IgG fucosylation in this study, possibly due to the power loss that accompanies a mathematically constructed trait of IgG_F (proportion of core-fucosylated species in the IgG N-glycome).

The limitations that we have encountered trying to estimate the effect of IgG N-glycan traits on disease risk relate to the generally lower power of the IgG N-glycome GWAS compared to the GWAS for the diseases. The maximum number of IVs per glycan trait that we were able to obtain was 9. New GWAS that involve larger cohorts and identify new genetic variants that are associated with IgG N-glycosylation will help overcoming this problem. We estimated the expected sample size for an IgG N-glycome GWAS to be able to detect the MR signal (glycans as exposure) if it exists. Median expected sample size was 25 500 and mean expected sample size was 30 350, with the maximum of 101 500 for Parkinson's disease (**Supplementary Table S8**). Given that new bigger GWASs for outcomes will be available in the future, it is reasonable to repeat the analysis when N-glycome GWAS of N~25 500/2=12 750 is available. (We assume here that the sample size for an outcome GWAS will be twice as high as that used in the current study)(58).

One of the important assumptions that must hold for Two-Sample MR analysis to be valid is the absence of confounders influencing both exposure and outcome. The set of genetic variants we used to test the causal effect of SLE on bisection of IgG glycans had no significant association with IgG_B on genome-wide level. The known risk factors for SLE include female sex, ageing, exogenous oestrogen administration, exposure to sun, vitamin D deficiency and consumption of certain drugs and exposure to some other biologically active substances, as pesticides, etc (59). There are differences in the IgG N-glycome associated with sex and age, however, the summary statistics that we used in our study was already corrected for sex and age. Oestrogen has a recorded impact on galactosylation of IgG N-glycans, however, no effect on bisection was reported (60). As for the other SLE risk factors, unfortunately, there have been no studies to our knowledge regarding their possible effect on IgG glycans. In addition, in this work we have not considered some possible confounders that are known to affect IgG N-glycans and could influence the risk of SLE. For instance, in most autoimmune diseases with pathogenic autoantibodies, as RA (4,61), SLE and others (62), overall N-glycosylation of the Fab (fragment antigen binding) portion of IgG is

increased. Fab N-glycome is known to contain more bisected species (63,64) and its possible expansion in other autoimmune diseases such as SLE potentially represents a confounder. One may speculate that pre-SLE or SLE status in IgG glycosylation cohorts could be a confounder too. However, given the low population prevalence of SLE (20 to 150 cases per 100,000 in USA)(65) the probability of inducing the false positive MR signal due to pre-SLE/SLE status is not likely. Future studies, analysing Fc and Fab N-glycomes of antigen-specific pathogenic antibodies in autoimmune diseases are expected to provide more information on this subject. All in all, any following research that will focus on the relationship between diseases and IgG N-glycome composition would benefit from studying potential confounding factors, not only connected with genetics, but also with lifestyle, effects of medical treatment, and other external influences.

To sum up, we performed a first large screening of possible causal relationships between the risk of 12 diseases and IgG N-glycosylation traits using the Two-Sample MR approach. We were unable to detect any significant effect of any of the IgG N-glycosylation traits on the risk of any diseases. We conclude that larger GWAS studies of IgG N-glycosylation are required to identify more genetic instruments needed to perform powerful MR studies with IgG N-glycosylation as an exposure. On the other hand, our data suggests that increased risk of SLE is causing increase in the relative abundance of N-glycan structures with bisecting GlcNAc in the total IgG glycome. According to our study and in agreement with the experimental data, elevated bisection seems to be downstream from the autoimmune disease, therefore we suggest this IgG glycosylation trait could be regarded as a biomarker of SLE.

Materials and Methods

Ethics statement

All studies providing genotype and phenotype data were approved by local Ethics Committees/Institutional Review Boards. Ethics approvals were given in compliance with the principles expressed in the Declaration of Helsinki. All study participants have signed written informed consent.

Participating cohorts

We selected 12 diseases (**Table 3**) to be included in the analysis based on the following criteria: 1) change of IgG glycosylation in the disease observed in a large scale glycomic case-control study (25); 2) publicly available summary statistics for the GWAS of the trait was available through the interactive database PheLiGe (49) at the moment when the current study was performed (December 2019). Some of the selected diseases shared genetic associations with the IgG N-glycosylation traits (36). All the summary statistics used are related to the cohorts of European descent. In case of CAD, CD, IBD, RA, T2D, UC for which the corresponding published GWAS are trans-ancestry, we only used the summary statistics related to the meta-analysis of European cohorts.

Summary statistics for IgG N-glycosylation traits were obtained from the GWAS of the following cohorts of European descent (**Table 4**): 1) the designated 8K GWAMA (N=8090) comprising cohorts CROATIA-Korcula, CROATIA-Vis, ORCADES, TwinsUK (dataset available at https://datashare.ed.ac.uk/handle/10283/3238) (66); 2) FINRISK (N=552)(36); EGCUT (N=575)(36); CROATIA-Korcula2 (N=941)(67); VIKING (N=1079)(68). Description of the CROATIA-Korcula, CROATIA-Vis, ORCADES, TwinsUK, EGCUT and FINRISK cohorts as well as details on phenotyping- and genotyping-related procedures can be found in (36). The Viking Health Study - Shetland (VIKING) is a family-based, cross-sectional study that seeks to identify genetic factors influencing cardiovascular and other disease risk in the population isolate of the Shetland Isles in northern Scotland (68). Genetic diversity in this population is decreased compared to Mainland Scotland, consistent with the high levels of endogamy historically. 2105 participants were recruited between 2013 and 2015, most having at least three grandparents from Shetland. Fasting blood samples were collected, and many health-related phenotypes and environmental exposures were measured in each individual. All participants gave informed consent, and the study was approved by the South East Scotland Research Ethics Committee, NHS Lothian (reference: 12/SS/0151). The CROATIA-Korcula2 study is a population based study that sampled participants, aged 18-92 years old, from the Adriatic island of Korčula in 2012 as an extension of the 10001 Dalmatians Biobank(67).

In each of the above mentioned cohorts GWAS were performed for 23 N-glycosylation traits (IGP1-23, **Supplementary Table S1 A**), that correspond to the relative abundances of the 23 chromatographic peaks in the UHPLC analysis of released IgG N-glycans (36). Each peak contains one major N-glycan structure(11) mainly representing diantennary complex N-glycans, with or without such features as core-fucose, bisecting *N*-acetylglycosamine (GlcNAc), terminal galactose

and terminal sialic acids on the antennae (**Fig. 1**). In the 8K GWAMA, as well as in FINRISK and EGCUT cohorts additional 54 derived N-glycosylation traits were measured. Derived traits describe abundances of specific groups of glycans, which share some structural characteristics, or their ratios (IGP24-77, **Supplementary Table S1 A).** Detailed description of the CROATIA-Korcula, CROATIA-Vis, ORCADES, TwinsUK, FINRISK and EGCUT cohorts, genotyping, quality control, imputation and GWAS can be found in (36,37).

Meta-analysis

Meta-analysis of the IgG N-glycosylation GWAS for the four cohorts: CROATIA-Korcula, CROATIA-Vis, ORCADES, TwinsUK (N total=8090) was performed in (36) and can be downloaded from https://datashare.ed.ac.uk/handle/10283/3238(66). The resulting summary association statistics in these samples will be referred to as the 8K GWAMA. In addition, we performed fixed effect inverse-variance meta-analysis, combining GWAS results in the following cohorts: a) FINRISK and EGCUT, (N total = 1127), later referred to as 1K GWAMA; b) 8K, FINRISK and EGCUT (N total = 9217), 9K GWAMA; c) FINRISK, EGCUT, VIKING and CROATIA-Korcula2 (N total = 3147), 3K GWAMA (see **Table 4**).

Compound derived traits

Most general traits, such as relative abundances of all core-fucosylated, galactosylated, sialylated, bisected N-glycans, were not included in the list of derived traits IGP1-77 and, therefore, no GWAS summary statistics were available for them. Nevertheless, to include these potentially informative traits we have constructed summary statistics using the approach described in (69) for the 9 general traits: fucosylation, bisection, agalactosylation, galactosylation, monogalactosylation, digalactosylation, sialylation, monosialylation and disialylation. GWAS summary statistic for each new compound trait was recalculated based on the GWAS summary statistics for the glycan traits IGP1-23 that contribute to the new trait (**Supplementary Table S1 B**).

As transformation coefficients we have used standard deviations (SD) of each IGP measurement in the CROATIA-Korcula cohort. Prior to SD estimation, N-glycosylation measurements were corrected for sex, age and kinship with the function “polygenic” from the GenABEL R package. Normalization to the SD was necessary because for the GWA analysis all glycosylation phenotypes were transformed to normal distribution with mean at 0 and SD 1(34), however, the non-

transformed glycan measurements have different standard deviations, dependent on the biological variation as well as on the precision of measurement of each glycan peak(70).

In brief, new effects and corresponding standard errors were calculated as in (69):

$$\beta_{new} = \frac{B*a}{\sqrt{sum(\Sigma_{ph}*(a \otimes a))}},$$

$$se_{new} = \frac{\sqrt{\left(se_1^2 + \frac{\beta_1^2}{N}\right) * sum\left(\Sigma_{ph} * (a \otimes a)\right) - \frac{(B*a)^2}{N}}}{\sqrt{sum(\Sigma_{ph} * (a \otimes a))}}$$

B – matrix, where each column β_i – effect of each SNP in $GWAS_i$; a – vector of transformation coefficients; Σ_{ph} – matrix of phenotypic correlation between GWAS; se_1 – se of each SNP's effect in 1st GWAS (corresponding to β_1); N – minimal sample size out of all GWAS for each SNP.

Two Sample MR analysis

We performed Two Sample MR using the "TwoSampleMR" R package (39) and its implementation in the framework of the GWAS-MAP platform (49,71). The causal effect estimate was obtained through an inverse variance weighted meta-analysis of ratios of exposure effect size and outcome effect size for each of the instruments. In the discovery analysis we were testing causal effects in all possible pairs of IgG N-glycosylation traits and diseases in both directions. For the two signals, detected in the discovery round, namely, a) causal effect of IGP76 on IBD and b) causal effect of SLE on IgG_B, we performed sensitivity analyses, univariate) and multivariate MR replication analyses (only for the effect of SLE on IgG_B), as described below, see also **Fig. 2.**

Discovery round of MR

Causal effect of IgG N-glycome on diseases

The procedure of the IV selection to estimate causal effects of IgG N-glycosylation traits on the 12 diseases was designed to ensure usage of genetic variants robustly associated with the exposure traits. We selected 31 genetic variants that were reported to be significantly associated with IgG N-glycosylation traits at genome-wide level (p-value of association $\leq 5 \times 10^{-8}$) in previous GWAS

of IgG N-glycosylation and for which this association was replicated independently (34–37) (**Supplementary Table S9**). Three of these SNPs, rs116108880, rs35590487 and rs12019136, were absent from at least one of our study cohorts. For rs116108880 we found a proxy SNPs using LDProxy Tool (https://ldlink.nci.nih.gov/?tab=home) linked with initial SNPs with $R^2 > 0.9$. For the other two genetic variants no suitable proxy SNPs were available. Thus, the final set of IVs comprised 29 genetic variants that are significantly associated with at least one individual IgG N-glycosylation trait.

GWAS summary statistics was available for 86 IgG N-glycosylation traits in two cohorts, 8K and 1K. To ensure that for each of the individual IgG N-glycosylation traits we are selecting genetic variants that are truly associated with each individual trait, for each of the 29 SNPs we checked if the following conditions are satisfied in the 8K and 1K GWAMA: 1) associated $p \leq 5x10^{-5}$(72) in the 8K GWAMA and associated $p \leq 0.05$ in 1K GWAMA; b) the associated effects of the SNP on the trait are directionally consistent between 8K and 1K GWAMAs. In this way we defined smaller subsets of IVs robustly associated with each individual IgG N-glycosylation trait. For the Two-Sample MR analysis of the IgG N-glycosylation effect on the 12 diseases we selected the 72 IgG N-glycosylation traits for which at least two IVs could be defined (**Supplementary Table S2)**. based on the conditions described above.

To use the largest and most powerful GWAS summary statistics as exposure for the IgG N-glycosylation traits, we performed meta-analysis of the 8K and 1K cohorts for the 86 IgG N-glycosylation traits and we refer to this meta-analysis as to the 9K GWAMA. We have calculated F-statistics and proportion of variance explained in each of the IgG N-glycosylation traits by the genetic variants chosen as IVs for MR analyses as a measure of the strength of the instruments (see **Supplementary Table S2**).

As outcomes we used publicly available summary statistics for the 12 diseases, described in detail in **Table 3**. The Bonferroni-corrected significance threshold of $p \leq 1.98 \times 10^{-4}$ was used, (0.05/21 principal components of IgG N-glycosylation traits *12 diseases).

Causal effect of diseases on IgG N-glycome

To perform a discovery round of MR for causal effect of disease risk on IgG N-glycosylation, from the GWAS of the corresponding disease we selected the SNPs that were associated with the disease at the genome-wide level ($p \leq 5 \times 10^{-8}$) with MAF ≥ 0.05. Then we performed clumping of the selected SNPs within a window of 500 Kb to obtain a set of independent IVs. As outcomes we used summary statistics of 86 IgG N-glycosylation traits (IGP1-77 and 9 compound traits) in the 8K

meta-analysis. The Bonferroni-corrected significance threshold of $p \leq 1.98x10^{-4}$ was used, (0.05/21 principal components of IgG N-glycosylation traits *12 diseases).

Sensitivity analysis of the causal effect of IGP76 on IBD

For the causal effect of IGP76 on IBD with sub-threshold statistical significance we identified outlier instrumental variables from both 'very good' and 'good' SNP sets corresponding to this pair of traits (p-values of the test for outliers in MR-PRESSO <1) and performed MR excluding those instrumental variables (**Supplementary Table S4**). As outcome the same GWAS of IBD (41) as in the discovery round was used.

Sensitivity analysis of the causal effect of SLE on bisection of IgG N-glycans

As Sensitivity analysis of the causal effect of SLE on IgG_B we performed TwoSample MR, using as instruments a refined set of 41 SLE-associated autosomal SNPs, considered replicated in(45) and summary statistics in the 8K GWAMA for bisection of IgG N-glycans (IgG_B) as an outcome. When a SNP from this list was not present in the outcome GWAS, we used a proxy SNP, defined as a SNP present in the outcome GWAS with the lowest associated p-value within 500 Kb window from the SNP of interest (**Supplementary Table S7**). We then used the MR-PRESSO R package to test if any of the instruments were outliers; and removed 5 instruments for which p-values of test for outliers in MR-PRESSO were <1 and repeated the analysis with 36 remaining instruments (**Supplementary Table S7**).

Replication of the causal effect of SLE on bisection of IgG N-glycans

To replicate the causal effect of SLE on bisection of IgG N-glycans (IgG_B), as instruments for Two-Sample MR we used the set of 36 genetic variants without outliers associated with SLE which was also used for the Sensitivity analysis (**Supplementary Table S7**) and corresponding summary statistics from(45). As the outcome we used summary statistics for IgG_B in the 3K GWAMA.

Multivariate MR of the SLE effect on IgG_B

To further replicate the most promising finding of a causal effect of SLE risk on IgG bisection we performed MV MR analysis (detailed pipeline in **Supplementary Fig. S7**), following the method described in (52). Calculations were performed using the MultiABEL R package (version 1.1-9, "MultiSummary" function) (35,52).

Using the method described in (73) we calculated the phenotypic correlations between the N-glycosylation traits (**Supplementary Fig. S6)** for the 8K GWAMA. Then we selected those of the IGP1-23 N-glycosylation traits that were strongly ($|r| > 0.2$) correlated with IgG_B (see **Supplementary Table 10, Supplementary Fig. S6**). N-glycosylation traits IGP1-23 represent the directly measured relative abundances of chromatographic peaks, these measurements are available for all IgG N-glycosylation cohorts used in this study.

Discovery Stage of Multivariate Analysis of Variance (MANOVA) was performed for the univariate Two-Sample MR estimates of causal effects of SLE on the selected IGPs correlated with IgG_B obtained using 8K GWAMA summary statistics for the IGPs.

For MV MR Replication Stage 1 MANOVA was performed for the univariate Two-Sample MR estimates of causal effects of SLE on the selected IGPs correlated with IgG_B obtained using 3K GWAMA summary statistics for the IGPs (**Supplementary Table S11**).

For MV MR Replication Stage 2 we extracted the coefficients of linear combinations estimated in the MV MR Discovery (**Supplementary Table S12**). The coefficients were then used to perform MANOVA for the univariate Two-Sample MR estimates of causal effects of SLE on the selected IGPs correlated with IgG_B obtained using 3K GWAMA summary statistics for the IGPs

The causal effect was considered replicated in the MV MR if the following conditions have been met: 1) MANOVA test p-value in the MV MR Discovery stage passed the threshold p-value $\leq 1.98 \times 10^{-4}$ defined for the univariate MR analysis; 2) Replication Stage 1 and Stage 2 MANOVA p-values both were significant at 0.05; 4) causal effect estimates in Discovery and Stage 2 MV MR were directionally concordant.

Estimation of expected sample size for N-glycome GWAS for detection of significant MR signal

The power of MR analysis depends only on the standard error associated with the effect of IVs in the outcome, so we can assume that the MR power (non-centrality parameter, NCP) depends

linearly on the number of Ivs. Number of Ivs is linearly dependent on the sample size of the exposure (58). We assumed the strongest (by p-value) MR signal for each of twelve diseases a true negative (**Supplementary Table S3**) and the sample size of discovery cohort (exposure), where these IVs were detected, as 8000. Then we estimated the upper limit for the expected sample size and the expected number of IVs to achieve the power of detection of an MR signal equal or more than 80% with $\leq 1.98 \times 10^{-4}$ (our discovery threshold) as following:

1) For each outcome (disease) we estimated the discovery NCP of casual effect as chi-square with df=1 from minimal p-value among all IGPs (*NCP_disc*). The corresponding number of IVs was used as a discovery number of IVs for each corresponding outcome (*N_IVs_disc*).

2) We estimated the expected number of IVs for the exposure as linear function of the exposure sample size as *N_IVs_expected= N_IVs_disc*N_new_sample_size/8000.*

3) The expected NCP was estimated as *NCP_expected= NCP_disc* N_IVs_expected/N_IVs_disc.*

4) We reported the minimal expected sample size needed to achieve 80% power of MR effect detection with $p \leq 1.98 \times 10^{-4}$ using R command "*pchisq(qchisq(1.98e-4, df=1, low=F), df=1, low=F, ncp= NCP_expected)*".

Data availability

Data obtained in the analyses are provided in Supplemental Tables related to this article.

Funding

The work of OOZ and GL was supported by the Croatian National Centre of Research Excellence in Personalized Healthcare grant (#KK.01.1.1.01.0010). The work of SZSh and YSA was funded by the Russian Science Foundation grant number 19-15-00115. The work of YAT was supported by the Ministry of Education and Science of the RF via the Institute of Cytology and Genetics (project 0259-2021-0009 / AAAA-A17-117092070032-4) and by the Russian Ministry of Science and Education under the 5-100 Excellence Programme. The work of AL was funded by the European Union's Horizon 2020 research and innovation program IMforFUTURE, under H2020-MSCA-ITN grant agreement number 721815. We acknowledge support from the MRC Human Genetics Unit programme grant, "Quantitative traits in health and disease" (U. MC_UU_00007/10).

Acknowledgments

The Viking Health Study – Shetland (VIKING) was supported by the MRC Human Genetics Unit quinquennial programme grant "QTL in Health and Disease". DNA extractions and genotyping were performed at the Edinburgh Clinical Research Facility, University of Edinburgh. We would like to acknowledge the invaluable contributions of the research nurses in Shetland, the administrative team in Edinburgh and the people of Shetland.

The Orkney Complex Disease Study (ORCADES) was supported by the Chief Scientist Office of the Scottish Government (CZB/4/276, CZB/4/710), a Royal Society URF to J.F.W., the MRC Human Genetics Unit quinquennial programme "QTL in Health and Disease", Arthritis Research UK and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947). DNA extractions were performed at the Edinburgh Clinical Research Facility, University of Edinburgh. We would like to acknowledge the invaluable contributions of the research nurses in Orkney, the administrative team in Edinburgh and the people of Orkney.

We would like to acknowledge the contribution of Dr. Tatiana I. Shashkova, Viktoriya Voroshilova and Dr. Anatolii Kirichenko who did an excellent job of maintaining the GWAS-MAP database and advising on how to use it in a most efficient way.

Conflicts of Interest Statement

GL is a founder and owner of the Genos Glycoscience Research Laboratory, and OOZ is an employee of the said company. YSA is a cofounder and a co-owner of PolyOmica and PolyKnomics, private organizations providing services, research, and development in the field of computational and statistical genomics. The other authors declare that they have no competing interests.

References

1. Juszczak, A., Pavić, T., Vučković, F., Bennett, A.J., Shah, N., Medvidović, E.P., Groves, C.J., Šekerija, M., Chandler, K., Burrows, C., *et al.* (2019) Plasma Fucosylated Glycans and C-Reactive Protein as Biomarkers of HNF1A-MODY in Young Adult–Onset Nonautoimmune Diabetes. *Diabetes Care*, **42**, 17–26.
2. Wittenbecher, C., Štambuk, T., Kuxhaus, O., Rudman, N., Vučković, F., Štambuk, J., Schiborn, C., Rahelić, D., Dietrich, S., Gornik, O., *et al.* (2020) Plasma N-Glycans as Emerging Biomarkers of Cardiometabolic Risk: A Prospective Investigation in the EPIC-Potsdam Cohort Study.

Diabetes Care, **43**, 661–668.
3. Verhelst, X., Dias, A.M., Colombel, J.-F., Vermeire, S., Van Vlierberghe, H., Callewaert, N. and Pinho, S.S. (2020) Protein Glycosylation as a Diagnostic and Prognostic Marker of Chronic Inflammatory Gastrointestinal and Liver Diseases. *Gastroenterology*, **158**, 95–110.
4. Hafkenscheid, L., Moel, E., Smolik, I., Tanner, S., Meng, X., Jansen, B.C., Bondt, A., Wuhrer, M., Huizinga, T.W.J., Toes, R.E.M., *et al.* (2019) *N* -Linked Glycans in the Variable Domain of IgG Anti–Citrullinated Protein Antibodies Predict the Development of Rheumatoid Arthritis. *Arthritis Rheumatol.*, **71**, 1626–1633.
5. Scott, D.A., Norris-Caneda, K., Spruill, L., Bruner, E., Kono, Y., Angel, P.M., Mehta, A.S. and Drake, R.R. (2019) Specific N-linked glycosylation patterns in areas of necrosis in tumor tissues. *Int. J. Mass Spectrom.*, **437**, 69–76.
6. Krištić, J., Vučković, F., Menni, C., Klarić, L., Keser, T., Beceheli, I., Pučić-Baković, M., Novokmet, M., Mangino, M., Thaqi, K., *et al.* (2014) Glycans Are a Novel Biomarker of Chronological and Biological Ages. *Journals Gerontol. Ser. A*, **69**, 779–789.
7. Vilaj, M., Gudelj, I., Trbojević-Akmačić, I., Lauc, G. and Pezer, M. (2019) IgG Glycans as a Biomarker of Biological Age. In Moskalev, A. (ed.), *Biomarkers of Human Aging. Healthy Ageing and Longevity, vol 10.*, Springer, Cham, pp. 81–99.
8. Opdenakker, G., Rudd, P.M., Ponting, C.P., Dwek, R.A. and Rudd, M. (1993) Concepts and principles of glycobiology. *FASEB J.*, **7**, 1330–7.
9. Varki, A. and Gagneux, P. (2017) Biological Functions of Glycans. In Varki, A., Cummings, R., Esko, J., Stanley, P., Hart, G., Aebi, M., Darvill, A., Kinoshita, T., Packer, N., Prestegard, J., et al. (eds.), *Essentials of Glycobiology [Online]*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).
10. Clerc, F., Reiding, K.R., Jansen, B.C., Kammeijer, G.S.M., Bondt, A. and Wuhrer, M. (2016) Human plasma protein N-glycosylation. *Glycoconj. J.*, **33**, 309–343.
11. Pucic, M., Knezevic, A., Vidic, J., Adamczyk, B., Novokmet, M., Polasek, O., Gornik, O., Supraha-Goreta, S., Wormald, M.R., Redzic, I., *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-M111.010090.
12. Ruhaak, L.R., Kim, K., Stroble, C., Taylor, S.L., Hong, Q., Miyamoto, S., Lebrilla, C.B. and Leiserowitz, G. (2016) Protein-Specific Differential Glycosylation of Immunoglobulins in Serum of Ovarian Cancer Patients. *J. Proteome Res.*, **15**, 1002–1010.
13. Kazuno, S., Furukawa, J., Shinohara, Y., Murayama, K., Fujime, M., Ueno, T. and Fujimura, T. (2016) Glycosylation status of serum immunoglobulin G in patients with prostate diseases. *Cancer Med.*, **5**, 1137–1146.
14. Kawaguchi-Sakita, N., Kaneshiro-Nakagawa, K., Kawashima, M., Sugimoto, M., Tokiwa, M., Suzuki, E., Kajihara, S., Fujita, Y., Iwamoto, S., Tanaka, K., *et al.* (2016) Serum immunoglobulin G Fc region N-glycosylation profiling by matrix-assisted laser desorption/ionization mass spectrometry can distinguish breast cancer patients from cancer-free controls. *Biochem. Biophys. Res. Commun.*, **469**, 1140–1145.
15. Vučković, F., Theodoratou, E., Thaci, K., Timofeeva, M., Vojta, A., Stambuk, J., Pucic-Bakovic, M., Rudd, P.M., Erek, L., Servis, D., *et al.* (2016) IgG Glycome in Colorectal Cancer. *Clin. Cancer Res.*, **22**, 1–10.
16. Chen, G., Wang, Y., Qin, X., Li, H., Guo, Y., Wang, Y., Liu, H., Wang, X., Song, G., Li, F., *et al.* (2013) Change in IgG $_1$ Fc *N* -linked glycosylation in human lung cancer: Age- and sex-related diagnostic potential. *Electrophoresis*, **34**, 2407–2416.
17. Parekh, R., Isenberg, D., Ansell, B., Roitt, I., Dwek, R. and Rademacher, T. (1988) Galactosylation Of Igg Associated Oligosaccharides: Reduction In Patients With Adult And Juvenile Onset Rheumatoid Arthritis And Relation To Disease Activity. *Lancet*, **331**, 966–969.

18. Vučković, F., Krištić, J., Gudelj, I., Teruel, M., Keser, T., Pezer, M., Pučić-Baković, M., Štambuk, J., Trbojevïc-Akmačić, I., Barrios, C., *et al.* (2015) Association of systemic lupus erythematosus with decreased immunosuppressive potential of the IgG glycome. *Arthritis Rheumatol.*, **67**, 2978–2989.
19. Šimurina, M., de Haan, N., Vučković, F., Kennedy, N.A., Štambuk, J., Falck, D., Trbojević-Akmačić, I., Clerc, F., Razdorov, G., Khon, A., *et al.* (2018) Glycosylation of Immunoglobulin G Associates With Clinical Features of Inflammatory Bowel Diseases. *Gastroenterology*, **154**, 1320-1333.e10.
20. Russell, A.C., Šimurina, M., Garcia, M.T., Novokmet, M., Wang, Y., Rudan, I., Campbell, H., Lauc, G., Thomas, M.G. and Wang, W. (2017) The N-glycosylation of immunoglobulin G as a novel biomarker of Parkinson's disease. *Glycobiology*, **27**, 501–510.
21. Lundström, S.L., Yang, H., Lyutvinskiy, Y., Rutishauser, D., Herukka, S.-K., Soininen, H. and Zubarev, R.A. (2013) Blood Plasma IgG Fc Glycans are Significantly Altered in Alzheimer's Disease and Progressive Mild Cognitive Impairment. *J. Alzheimer's Dis.*, **38**, 567–579.
22. Bermingham, M.L., Colombo, M., McGurnaghan, S.J., Blackbourn, L.A.K., Vučković, F., Pučić Baković, M., Trbojević-Akmačić, I., Lauc, G., Agakov, F., Agakova, A.S., *et al.* (2018) N-Glycan Profile and Kidney Disease in Type 1 Diabetes. *Diabetes Care*, **41**, 79–87.
23. Lemmers, R.F.H., Vilaj, M., Urda, D., Agakov, F., Šimurina, M., Klaric, L., Rudan, I., Campbell, H., Hayward, C., Wilson, J.F., *et al.* (2017) IgG glycan patterns are associated with type 2 diabetes in independent European populations. *Biochim. Biophys. Acta - Gen. Subj.*, **1861**, 2240–2249.
24. Wang, Y., Klarić, L., Yu, X., Thaqi, K., Dong, J., Novokmet, M., Wilson, J., Polasek, O., Liu, Y., Krištić, J., *et al.* (2016) The Association between Glycosylation of Immunoglobulin G and Hypertension. *Med. (United States)*, **95**.
25. Gudelj, I., Lauc, G. and Pezer, M. (2018) Immunoglobulin G glycosylation in aging and diseases. *Cell. Immunol.*, **333**, 65–79.
26. Nimmerjahn, F., Gordan, S. and Lux, A. (2015) FcγR dependent mechanisms of cytotoxic, agonistic, and neutralizing antibody activities. *Trends Immunol.*, **36**, 325–336.
27. Nimmerjahn, F. and Ravetch, J. V (2005) Divergent immunoglobulin g subclass activity through selective Fc receptor binding. *Science*, **310**, 1510–2.
28. Schwab, I., Mihai, S., Seeling, M., Kasperkiewicz, M., Ludwig, R.J. and Nimmerjahn, F. (2014) Broad requirement for terminal sialic acid residues and FcγRIIB for the preventive and therapeutic activity of intravenous immunoglobulins in vivo. *Eur. J. Immunol.*, **44**, 1444–1453.
29. Quast, I., Keller, C.W., Maurer, M.A., Giddens, J.P., Tackenberg, B., Wang, L.X., Münz, C., Nimmerjahn, F., Dalakas, M.C. and Lünemann, J.D. (2015) Sialylation of IgG Fc domain impairs complement-dependent cytotoxicity. *J. Clin. Invest.*, **125**, 4160–4170.
30. Rombouts, Y., Ewing, E., Van De Stadt, L.A., Selman, M.H.J., Trouw, L.A., Deelder, A.M., Huizinga, T.W.J., Wuhrer, M., Van Schaardenburg, D., Toes, R.E.M., *et al.* (2015) Anti-citrullinated protein antibodies acquire a pro-inflammatory Fc glycosylation phenotype prior to the onset of rheumatoid arthritis. *Ann. Rheum. Dis.*, **74**, 234–241.
31. Kellokumpu, S. (2019) Golgi pH, ion and redox homeostasis: How much do they really matter? *Front. Cell Dev. Biol.*, **7**, 93.
32. Pothukuchi, P., Agliarulo, I., Russo, D., Rizzo, R., Russo, F. and Parashuraman, S. (2019) Translation of genome to glycome: role of the Golgi apparatus. *FEBS Lett.*, **593**, 2390–2411.
33. Ohtsubo, K. and Marth, J.D. (2006) Glycosylation in Cellular Mechanisms of Health and Disease. *Cell*, **126**, 855–867.
34. Lauc, G., Huffman, J.E., Pučić, M., Zgaga, L., Adamczyk, B., Mužinić, A., Novokmet, M., Polašek, O., Gornik, O., Krištić, J., *et al.* (2013) Loci Associated with N-Glycosylation of Human Immunoglobulin G Show Pleiotropy with Autoimmune Diseases and Haematological Cancers. *PLoS Genet.*, **9**, e1003225.

35. Shen, X., Klarić, L., Sharapov, S., Mangino, M., Ning, Z., Wu, D., Trbojević-Akmačić, I., Pučić-Baković, M., Rudan, I., Polašek, O., *et al.* (2017) Multivariate discovery and replication of five novel loci associated with Immunoglobulin G N-glycosylation. *Nat. Commun.*, **8**, 447.
36. Klarić, L., Tsepilov, Y.A., Stanton, C.M., Mangino, M., Sikka, T.T., Esko, T., Pakhomov, E., Salo, P., Deelen, J., McGurnaghan, S.J., *et al.* (2020) Glycosylation of immunoglobulin G is regulated by a large network of genes pleiotropic with inflammatory diseases. *Sci. Adv.*, **6**, eaax0301.
37. Shadrina, A.S., Zlobin, A.S., Zaytseva, O.O., Klaric, L., Sharapov, S.Z., Pakhomov, E., Perola, M., Esko, T., Hayward, C., Wilson, J.F., *et al.* (2021) Multivariate genome-wide analysis of immunoglobulin G N-glycosylation identifies new loci pleiotropic with immune function. *Hum. Mol. Genet.*
38. Lawlor, D.A., Harbord, R.M., Sterne, J.A.C., Timpson, N. and Davey Smith, G. (2008) Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat. Med.*, **27**, 1133–1163.
39. Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., *et al.* (2018) The MR-base platform supports systematic causal inference across the human phenome. *Elife*, **7**.
40. Teumer, A. (2018) Common Methods for Performing Mendelian Randomization. *Front. Cardiovasc. Med.*, **5**, 28.
41. Liu, J.Z., Van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., *et al.* (2015) Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.*, **47**, 979–986.
42. Mahajan, A., Go, M.J., Zhang, W., Below, J.E., Gaulton, K.J., Ferreira, T., Horikoshi, M., Johnson, A.D., Ng, M.C.Y., Prokopenko, I., *et al.* (2014) Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.*, **46**, 234–244.
43. Nikpay, M., Goel, A., Won, H.-H., Hall, L.M., Willenborg, C., Kanoni, S., Saleheen, D., Kyriakou, T., Nelson, C.P., Hopewell, J.C., *et al.* (2015) A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.*, **47**, 1121–1130.
44. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., *et al.* (2014) Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*, **506**, 376–381.
45. Bentham, J., Morris, D.L., Cunninghame Graham, D.S., Pinder, C.L., Tombleson, P., Behrens, T.W., Martín, J., Fairfax, B.P., Knight, J.C., Chen, L., *et al.* (2015) Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat. Genet.*, **47**, 1457–1464.
46. Jansen, I.E., Savage, J.E., Watanabe, K., Bryois, J., Williams, D.M., Steinberg, S., Sealock, J., Karlsson, I.K., Hägg, S., Athanasiu, L., *et al.* (2019) Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.*, **51**, 404–413.
47. Pankratz, N., Beecham, G.W., DeStefano, A.L., Dawson, T.M., Doheny, K.F., Factor, S.A., Hamza, T.H., Hung, A.Y., Hyman, B.T., Ivinson, A.J., *et al.* (2012) Meta-analysis of Parkinson's Disease: Identification of a novel locus, RIT2. *Ann. Neurol.*, **71**, 370–384.
48. Wang, Y., McKay, J.D., Rafnar, T., Wang, Z., Timofeeva, M.N., Broderick, P., Zong, X., Laplana, M., Wei, Y., Han, Y., *et al.* (2014) Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nat. Genet.*, **46**, 736–741.
49. Shashkova, T.I., Pakhomov, E.D., Gorev, D.D., Karssen, L.C., Joshi, P.K. and Aulchenko, Y.S. (2021) PheLiGe: an interactive database of billions of human genotype-phenotype associations. *Nucleic Acids Res.*, **49**, D1347–D1350.
50. Shashkova, T., Gorev, D., Pakhomov, E., Shadrina, A., Sharapov, Sz., Tsepilov, Y., Karssen, L., Aulchenko, Y., Шашкова, Т., Горев, Д., *et al.* (2020) The GWAS-MAP platform for aggregation

of results of genome-wide association studies and the GWAS-MAP|homo database of 70 billion genetic associations of human traits Платформа GWAS-MAP для агрегации результатов полногеномных исследований ассоциаций и база. *Vavilov J. Genet. Breed.*, **24**, 876–884.
51. Verbanck, M., Chen, C.-Y., Neale, B. and Do, R. (2018) Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.*, **50**, 693–698.
52. Ning, Z., Tsepilov, Y.A., Sharapov, S.Z., Wang, Z., Grishenko, A.K., Feng, X., Shirali, M., Joshi, P.K., Wilson, J.F., Pawitan, Y., *et al.* (2021) Nontrivial Replication of Loci Detected by Multi-Trait Methods. *Front. Genet.*, **12**.
53. Yarwood, A., Viatte, S., Okada, Y., Plenge, R., Yamamoto, K., Barton, A., Symmons, D., Raychaudhuri, S., Klareskog, L., Gregersen, P., *et al.* (2016) Loci associated with N-glycosylation of human IgG are not associated with rheumatoid arthritis: A Mendelian randomisation study. *Ann. Rheum. Dis.*, **75**, 317–320.
54. Trbojević Akmačić, I., Ventham, N.T., Theodoratou, E., Vučković, F., Kennedy, N.A., Krištić, J., Nimmo, E.R., Kalla, R., Drummond, H., Štambuk, J., *et al.* (2015) Inflammatory bowel disease associates with proinflammatory potential of the immunoglobulin G glycome. *Inflamm. Bowel Dis.*, **21**, 1237–47.
55. Umaña, P., Jean-Mairet, J., Moudry, R., Amstutz, H. and Bailey, J.E. (1999) Engineered glycoforms of an antineuroblastoma IgG1 with optimized antibody-dependent cellular cytotoxic activity. *Nat. Biotechnol.*, **17**, 176–180.
56. Shinkawa, T., Nakamura, K., Yamane, N., Shoji-Hosaka, E., Kanda, Y., Sakurada, M., Uchida, K., Anazawa, H., Satoh, M., Yamasaki, M., *et al.* (2003) The absence of fucose but not the presence of galactose or bisecting N-acetylglucosamine of human IgG1 complex-type oligosaccharides shows the critical role of enhancing antibody-dependent cellular cytotoxicity. *J. Biol. Chem.*, **278**, 3466–73.
57. Ikeda, Y., Ihara, H., Tsukamoto, H., Gu, J. and Taniguchi, N. (2014) Mannosyl (beta-1,4-)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase (MGAT3); β1,4-NAcetylglucosaminyltransferase III (GnT-III, GlcNAcT-III). *Handbook of Glycosyltransferases and Related Genes, Second Edition*, Springer Japan, Vol. 1, pp. 209–222.
58. Canela-Xandri, O., Rawlik, K. and Tenesa, A. (2018) An atlas of genetic associations in UK Biobank. *Nat. Genet.*, **50**, 1593–1599.
59. Ceroni, A., Maass, K., Geyer, H., Geyer, R., Dell, A. and Haslam, S.M. (2008) GlycoWorkbench: A tool for the computer-assisted annotation of mass spectra of glycans. *J. Proteome Res.*, **7**, 1650–1659.
60. Ercan, A., Kohrt, W.M., Cui, J., Deane, K.D., Pezer, M., Yu, E.W., Hausmann, J.S., Campbell, H., Kaiser, U.B., Rudd, P.M., *et al.* (2017) Estrogens regulate glycosylation of IgG in women and men. *JCI insight*, **2**, e89703.
61. Bovenkamp, F.S. van de, Hafkenscheid, L., Rispens, T. and Rombouts, Y. (2016) The Emerging Importance of IgG Fab Glycosylation in Immunity. *J. Immunol.*, **196**, 1435–1441.
62. Visser, A., Hamza, N., Kroese, F.G.M. and Bos, N.A. (2018) Acquiring new N-glycosylation sites in variable regions of immunoglobulin genes by somatic hypermutation is a common feature of autoimmune diseases. Acquiring new N-glycosylation sites in variable regions of immunoglobulin genes by somatic hypermutation is a common feature of autoimmune diseases. *Ann. Rheum. Dis.* **(2018)** , *77*, e69.
63. Bondt, A., Rombouts, Y., Selman, M.H.J., Hensbergen, P.J., Reiding, K.R., Hazes, J.M.W., Dolhain, R.J.E.M. and Wuhrer, M. (2014) Immunoglobulin G (IgG) Fab Glycosylation Analysis Using a New Mass Spectrometric High-throughput Profiling Method Reveals Pregnancy-associated Changes. *Mol. Cell. Proteomics*, **13**, 3029–3039.
64. Hafkenscheid, L., Bondt, A., Scherer, H.U., Huizinga, T.W.J., Wuhrer, M., Toes, R.E.M. and

Rombouts, Y. (2017) Structural analysis of variable domain glycosylation of anti-citrullinated protein antibodies in rheumatoid arthritis reveals the presence of highly sialylated glycans. *Mol. Cell. Proteomics*, **16**, 278–287.
65. Lawrence, R.C., Helmick, C.G., Arnett, F.C., Deyo, R.A., Felson, D.T., Giannini, E.H., Heyse, S.P., Hirsch, R., Hochberg, M.C., Hunder, G.G., *et al.* (1998) Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. *Arthritis Rheum.*, **41**, 778–799.
66. Klaric, L., Tsepilov, Y.A., Aulchenko, Y.S., Lauc, G. and Hayward, C. (2019) GWAS summary statistics for UPLC IgG N-glycosylation traits. GWAS summary statistics for UPLC IgG N-glycosylation traits https://datashare.ed.ac.uk/handle/10283/3238 (accessed Feb 11, 2021).
67. Rudan, I., Marušić, A., Janković, S., Rotim, K., Boban, M., Lauc, G., Grković, I., Dogaš, Z., Zemunik, T., Vatavuk, Z., *et al.* (2009) "10 001 Dalmatians:" Croatia Launches Its National Biobank. "10 001 Dalmatians:" Croatia Launches Its National Biobank. *Croat. Med. J.* **(2009)** , *50*, 4–6.
68. Kerr, S.M., Klaric, L., Halachev, M., Hayward, C., Boutin, T.S., Meynert, A.M., Semple, C.A., Tuiskula, A.M., Swan, H., Santoyo-Lopez, J., *et al.* (2019) An actionable KCNH2 Long QT Syndrome variant detected by sequence and haplotype analysis in a population research cohort. *Sci. Rep.*, **9**, 10964.
69. Tsepilov, Y.A., Freidin, M.B., Shadrina, A.S., Sharapov, S.Z., Elgaeva, E.E., Zundert, J. van, Karssen, L.C., Suri, P., Williams, F.M.K. and Aulchenko, Y.S. (2019) Analysis of genetically independent phenotypes identifies shared genetic factors associated with chronic musculoskeletal pain at different anatomic sites. *bioRxiv*, 810283.
70. Trbojević Akmačić, I., Ugrina, I., Štambuk, J., Gudelj, I., Vučković, F., Lauc, G. and Pučić-Baković, M. (2015) High-throughput glycomics: Optimization of sample preparation. *Biochem.*, **80**, 934–942.
71. Shashkova, T., Gorev, D., Pakhomov, E., Shadrina, A., Sharapov, Sz., Tsepilov, Y., Karssen, L., Aulchenko, Y., Шашкова, Т., Горев, Д., *et al.* (2020) The GWAS-MAP platform for aggregation of results of genome-wide association studies and the GWAS-MAP|homo database of 70 billion genetic associations of human traits Платформа GWAS-MAP для агрегации результатов полногеномных исследований ассоциаций и база. *Vavilov J. Genet. Breed.*, **24**, 876–884.
72. Burgess, S., Thompson, S.G. and CRP CHD Genetics Collaboration (2011) Avoiding bias from weak instruments in Mendelian randomization studies. *Int. J. Epidemiol.*, **40**, 755–64.
73. Stephens, M. (2013) A Unified Framework for Association Analysis with Multiple Related Phenotypes. *PLoS One*, **8**, 65245.
74. Gatz, M., Reynolds, C.A., Fratiglioni, L., Johansson, B., Mortimer, J.A., Berg, S., Fiske, A. and Pedersen, N.L. (2006) Role of Genes and Environments for Explaining Alzheimer Disease. *Arch. Gen. Psychiatry*, **63**, 168–174.
75. Thomsen, S.F., Van Der Sluis, S., Kyvik, K.O., Skytthe, A. and Backer, V. (2010) Estimates of asthma heritability in a large twin sample. *Clin. Exp. Allergy*, **40**, 1054–1061.
76. Zdravkovic, S., Wienke, A., Pedersen, N.L., Marenberg, M.E., Yashin, A.I. and De Faire, U. (2002) Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J. Intern. Med.*, **252**, 247–254.
77. Khera, A. V and Kathiresan, S. (2017) Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat. Rev. Genet.*, **18**, 331–344.
78. Chen, G.-B., Lee, S.H., Brion, M.-J.A., Montgomery, G.W., Wray, N.R., Radford-Smith, G.L. and Visscher, P.M. (2014) Estimation and partitioning of (co)heritability of inflammatory bowel disease from GWAS and immunochip data. *Hum. Mol. Genet.*, **23**, 4710–4720.
79. Kupper, N., Willemsen, G., Riese, H., Posthuma, D., Boomsma, D.I. and de Geus, E.J.C. (2005) Heritability of daytime ambulatory blood pressure in an extended twin design. *Hypertens.*

(Dallas, Tex. 1979), **45**, 80–5.
80. McCaffery, J.M., Papandonatos, G.D., Lyons, M.J. and Niaura, R. (2008) Educational attainment and the heritability of self-reported hypertension among male Vietnam-era twins. *Psychosom. Med.*, **70**, 781–6.
81. Lichtenstein, P., Holm, N. V, Verkasalo, P.K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A. and Hemminki, K. (2000) Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.*, **343**, 78–85.
82. Czene, K., Lichtenstein, P. and Hemminki, K. (2002) Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish family-cancer database. *Int. J. Cancer*, **99**, 260–266.
83. Mucci, L.A., Hjelmborg, J.B., Harris, J.R., Czene, K., Havelick, D.J., Scheike, T., Graff, R.E., Holst, K., Möller, S., Unger, R.H., *et al.* (2016) Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA*, **315**, 68.
84. Wirdefeldt, K., Gatz, M., Reynolds, C.A., Prescott, C.A. and Pedersen, N.L. (2011) Heritability of Parkinson disease in Swedish twins: a longitudinal study. *Neurobiol. Aging*, **32**, 1923.e1-1923.e8.
85. MacGregor, A.J., Snieder, H., Rigby, A.S., Koskenvuo, M., Kaprio, J., Aho, K. and Silman, A.J. (2000) Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum.*, **43**, 30–37.
86. Almgren, P., Lehtovirta, M., Isomaa, B., Sarelin, L., Taskinen, M.R., Lyssenko, V., Tuomi, T. and Groop, L. (2011) Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study. *Diabetologia*, **54**, 2811–2819.
87. Willemsen, G., Ward, K.J., Bell, C.G., Christensen, K., Bowden, J., Dalgård, C., Harris, J.R., Kaprio, J., Lyle, R., Magnusson, P.K.E., *et al.* (2015) The Concordance and Heritability of Type 2 Diabetes in 34,166 Twin Pairs From International Twin Registers: The Discordant Twin (DISCOTWIN) Consortium. *Twin Res. Hum. Genet.*, **18**, 762–771.
88. Chen, G.B., Lee, S.H., Brion, M.J.A., Montgomery, G.W., Wray, N.R., Radford-Smith, G.L. and Visscher, P.M. (2014) Estimation and partitioning of (co)heritability of inflammatory bowel disease from GWAS and immunochip data. *Hum. Mol. Genet.*, **23**, 4710–4720.

Tables

Table 1. Two-Sample MR analysis of causal effect of IGP76 on IBD risk.

	Discovery Round				Sensitivity Analysis			
Method	**Number of IVs**	**β^a**	**SE[b]**	**P-value[c]**	**Number of IVs**	**β^d**	**SE[e]**	**P-value[f]**
Inverse variance weighted	6	0.182	0.051	3.66×10^{-4}	4	0.137	0.060	0.023
MR Egger	6	0.037	0.172	0.840	4	0.065	0.179	0.753
Simple mode	6	0.235	0.100	0.064	4	0.183	0.099	0.162
Weighted median	6	0.194	0.064	0.003	4	0.137	0.069	0.048
Weighted mode	6	0.220	0.101	0.081	4	0.192	0.093	0.132

[a, d] – Causal effects (betas) of IGP76 on IBD risk.
[b, e] – Standard errors of the causal effect of IGP76 on IBD risk
[c, f] - P-values of the causal effect of IGP76 on IBD risk.

Table 2. Two-Sample MR analysis of SLE risk effect on bisection of IgG N-glycans (IgG_B).

Analysis round	Exposure	Outcome	IV selection criteria	Number of IVs	Method	β^a	SE[b]	P-value[c]
Discovery MR	SLE	IgG_B 8K	Independent, associated with SLE at $p \leq 5 \times 10^{-8}$	36	Inverse variance weighted	0.123	0.024	5.118×10^{-7}
					MR Egger	0.184	0.053	0.001
					Simple mode	0.099	0.061	0.112
					Weighted median	0.100	0.033	0.003
					Weighted mode	0.117	0.049	0.023
Sensitivity analysis	SLE	IgG_B 8K	Reported as associated with SLE in (45); ouliers removed	36	Inverse variance weighted	0.124	0.028	1.240×10^{-5}
					MR Egger	0.145	0.053	0.010
					Simple mode	0.052	0.081	0.529
					Weighted median	0.094	0.042	0.025
					Weighted mode	0.109	0.048	0.029
Univariate replication	SLE	IgG_B 3K	Reported as associated with SLE in (45); ouliers removed	36	Inverse variance weighted	-0.016	0.044	0.720
					MR Egger	0.002	0.088	0.985
					Simple mode	0.010	0.123	0.937
					Weighted median	-0.033	0.063	0.607
					Weighted mode	-0.026	0.077	0.735
Multivarite MR analysis	SLE	N-glycan traits correlated with IgG_B	Reported as associated with SLE in (45); ouliers removed	36	MV MR Discovery (8K)	0.008	0.003	9.43×10^{-4}
					MV MR Replication 1 (3K)	0.018	0.005	5.98×10^{-4}
					MV MR Replication 2 (3K with coefficients obtained in Discovery MV MR)	0.091	0.031	0.003

[a] – Causal effect (beta) of SLE on IgG_B.
[b] – Standard errors of the causal effect of SLE on IgG_B.
[c]- P-values of the causal effect of SLE on IgG_B.

Table 3. Description of GWAS of disease risk.

Trait	Colocalization with IgG N-glycosylation loci	Data source	N cases	Sample size	Number of variants	Reference	Reported % of variance explained[a]	Total heritability
Alzheimer's disease	no	GWAS-MAP	71880	455258	9862738	(46)	7.1% (3.9% without APOE locus)	60-80% (74)
Asthma	yes	mrbase.org	53257	462013	9851867	UKBB 2018, 6152#8: (http://www.nealelab.is)	16.8%[b]	60% (75)
Coronary artery disease	no	mrbase.org	60801	184305	9455779	(43)	37%	40 - 60% (76,77)
Crohn's disease	yes	mrbase.org	5956	20883	12276506	(41)	13.1%[c]	75% (78)
Hypertension	no	mrbase.org	87690	337159	10894596	UKBB 2017, (http://www.nealelab.is)	23.8%[b]	45 - 60%(79,80)
Inflammatory bowel disease	yes	mrbase.org	12882	34652	12716084	(41)	NA[d]	NA
Lung cancer	no	mrbase.org	11348	27209	8945893	(48)	NA	8 - 25%(81–83)
Parkinson's disease	yes	GWAS-MAP	4238	8477	2494599	(47)	NA	19 – 34% (84)
Rheumatoid arthritis	yes	mrbase.org	14361	58284	9739304	(44)	5.5%	53 – 65% (85)
Systemic lupus erythematosus	no	GWAS-MAP	4036	10995	8690139	(45)	15.3%	65% (45)
Type II diabetes	no	mrbase.org	26488	110452	2915012	(42)	NA	30-70% (86,87)
Ulcerative colitis	yes	mrbase.org	6968	27432	12255197	(41)	8.2%[c]	67% (88)

[a] – as reported in the study used as a source of summary statistics for the current analysis
[b] – as reported at the web page https://nealelab.github.io/UKBB_ldsc/h2_browser.html
[c] - data for the final meta-analysis, including non-European populations
[d] - NA – information not available in the corresponding publication

Table 4. Description of GWAMA of IgG N-glycosylation traits. GWAS performed in 9K cohort (8K + 1K) were used as outcomes for the MR analysis of IgG N-glycosylation traits on 12 diseases as the most powerful GWAS available for all 86 N-glycosylation traits; GWAS performed in 8K were used as exposures for the MR analysis of 12 diseases on IgG N-glycosylation traits, while 3K GWAS were reserved for replication of the associations observed in the discovery round.

Cohort	Sample size	IgG glycosylation traits measured			GWAMA			
		IGP1-23	IGP24-77	9 compound traits	8K	9K	1K	3K
CROATIA-Korcula	849	+	+	+	+	+		
CROATIA-Vis	802	+	+	+	+	+		
ORCADES	1960	+	+	+	+	+		
TWINSUK	4479	+	+	+	+	+		
EGCUT	575	+	+	+		+	+	+
FINRISK	552	+	+	+		+	+	+
VIKING	1079	+		+				+
CROATIA-Korcula2	941	+		+				+

Legends to Figures

Fig. 1. Functional consequences of differential IgG N-glycosylation and diseases associated with changes in the corresponding trait (25). More details in the text.

Fig. 2. Schematic representation of the analysis. More details in the text.

Abbreviations

AD - Alzheimer's disease
CAD - coronary artery disease
CD - Crohn's disease
GWAS – genome-wide association study
GWAMA – genome-wide association meta-analysis
IBD - inflammatory bowel disease
IgG – immunoglobulin G
IV -instrumental variable
MANOVA – Multivariate Analysis of Variance
MR – Mendelian randomization
MV – Multivariate
NCP – non-centrality parameter
PD - Parkinson's disease
RA - rheumatoid arthritis
SLE - systemic lupus erythematous
SNP – single nucleotide polymorphism
T2D - type 2 diabetes
UC - ulcerative colitis
UHPLC – ultra high-pressure liquid chromatography