University of Zagreb
FACULTY OF PHARMACY AND BIOCHEMISTRY

Toma Keser

INFLUENCE OF GENETIC AND ENVIRONMENTAL FACTORS ON *N*-GLYCOSYLATION OF IMMUNOGLOBULIN G AND TOTAL PLASMA PROTEINS DETERMINED BY TWIN STUDY

DOCTORAL THESIS

Zagreb, 2016.

Sveučilište u Zagrebu
FARMACEUTSKO-BIOKEMIJSKI FAKULTET

Toma Keser

UTJECAJ GENSKIH I OKOLIŠNIH ČIMBENIKA NA *N*-GLIKOZILACIJU IMUNOGLOBULINA G I UKUPNIH PLAZMATSKIH PROTEINA ODREĐEN STUDIJOM NA BLIZANCIMA

DOKTORSKI RAD

Zagreb, 2016.

University of Zagreb
FACULTY OF PHARMACY AND BIOCHEMISTRY

Toma Keser

INFLUENCE OF GENETIC AND ENVIRONMENTAL FACTORS ON *N*-GLYCOSYLATION OF IMMUNOGLOBULIN G AND TOTAL PLASMA PROTEINS DETERMINED BY TWIN STUDY

DOCTORAL THESIS

Supervisor: Prof. Gordan Lauc, PhD

Zagreb, 2016.

Sveučilište u Zagrebu
FARMACEUTSKO-BIOKEMIJSKI FAKULTET

Toma Keser

UTJECAJ GENSKIH I OKOLIŠNIH ČIMBENIKA NA *N*-GLIKOZILACIJU IMUNOGLOBULINA G I UKUPNIH PLAZMATSKIH PROTEINA ODREĐEN STUDIJOM NA BLIZANCIMA

DOKTORSKI RAD

Mentor: prof.dr.sc. Gordan Lauc

Zagreb, 2016.

The doctoral thesis was submitted to the Faculty Council of the Faculty of Pharmacy and Biochemistry, University of Zagreb in order to acquire a PhD degree in the area of Biomedicine and Health, the field of Pharmacy, the branch of Pharmacy.

The work presented in this doctoral thesis was performed at the Faculty of Pharmacy and Biochemistry, University of Zagreb, under supervision of Prof. Gordan Lauc, PhD., in collaboration with the Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. The research conducted at the Faculty was financed by FP7 INTEGRA-LIFE Project (contract #315997).

ZAHVALE / ACKNOWLEDGEMENTS

Ponajprije hvala mom mentoru prof. dr. sc. Gordanu Laucu koji mi je omogućio da ostvarim svoju veliku želju i uveo me u svijet znanosti te me naučio kako biti samostalan i snaći se u svim segmentima znanstvenog rada.

Posebno hvala izv. prof. dr. sc. Olgi Gornik koja me pratila u mom znanstvenom hodu od prvih koraka u studentskim danima, u izradi rada za Rektorovu nagradu pa sve do kraja izrade moga doktorskog rada. Olga, hvala ti na svakom savjetu i pomoći te na prekrasnoj i konstruktivnoj radnoj atmosferi u našem FBF gliko timu u kojemu ne manjka smijeha i vedrine, pa ni onda kada smo svi pretrpani poslom.

Velika hvala i svim ostalim djelatnicima Zavoda za biokemiju i molekularnu biologiju koji su mi uvijek bili potpora i pokazali razumijevanje kada je to bilo potrebno.

Hvala Ani i Maji koje su me uvele u svijet glikobiologije i naučile me mojim prvim metodama analize glikana i koje su mi kasnije bile super šefice, svaka na svoj jedinstven način.

Hvala i mojoj labos cimerici Tamari s kojom sam podijelio mnogo laboratorijskih uspjeha i neuspjeha, smijeha i uzdaha, sreće i očaja te svakodnevno raspravljao o svakom našem eksperimentu, radu, izvještaju, prezentaciji, pa i o idejama za ručak i buduća putovanja.

Hvala mom prijatelju Mislavu na svim laboratorijskim razgovorima o MS-u, političkim debatama u menzi na Veterini i filozofskim raspravama uz pivu.

Hvala Genadiju što mi je pomogao da produbim svoje znanje o MS-u te mi dao mnoge korisne analitičke savjete, naravno i zbog toga što je nesebično svakodnevno dijelio svoje baklave.

Velika hvala svima iz Genosova tima - cijeloj mladoj, veseloj i pametnoj ekipi na svim korisnim prijedlozima, konstruktivnim raspravama (ponekad i malo manje konstruktivnim ☺), posuđivanju kemikalija i opreme, napornim sastancima i zabavnim druženjima.

Many thanks to prof. Manfred Wuhrer and all his team for generously sharing their knowledge with me and helping me to get familiar with the broad mass spectrometry use in glycomics and glycoproteomics.

Od srca hvala mojoj kolegici, najboljoj prijateljici i supruzi Sabini što mi je uvijek bila čvrst oslonac i nepresušan izvor radosti i smijeha. Veselim se svakom našem zajedničkom trenutku. "*The minute I heard my first love story, I started looking for you, not knowing how blind that was. Lovers don't finally meet somewhere. They're in each other all along.*"

Hvala mojim roditeljima na ljubavi, sigurnosti i slobodi koje su mi uvijek pružali i kojih mi nikada nije nedostajalo. Hvala i mojoj braći na neslomljivoj slozi. Uz naš su zajednički smijeh svi problemi i teškoće postajali manje ozbiljni i manje opterećujući.

Hvala kumovima Ljubici i Feriju na bezuvjetnoj podršci kojom su znatno pridonijeli mom školovanju.

Beskrajno sam zahvalan svom učitelju fra Mladenu Hercegu koji me naučio koliko je važna prisutnost i pribranost u sadašnjem trenutku te predanost onom što trenutno radiš.

Stupnjevi

Kao što vene cvat i svaku mladost
smjenjuje starost, života stupanj svaki
i mudrost svaka i vrlina svaka
u svoje vrijeme cvate i ne traje.
Uvijek kad život zove mora srce,
na rastanak i nov početak spremno,
bez tuge, hrabro, novoj vezi prići.
jer svakom je početku svojstven neki čar
što štiti nas i pomaže nam živjeti.

Prostor za prostorom vedro valja proći,
ni s jednim se ne vezat zavičajno:
uske i sputane Duh nas Svijeta neće,
stupanj po stupanj diže nas i širi;
jer čim u jednom krugu smo se našli
i svikli se, klonuće već nam prijeti.
Tek onaj putnik spreman da se otkine,
naviku ruši što sve snage veže.
Možda će i u času smrti naše
poslat nas mlade novim prostorima;
ne prestaje nas zov života zvati:
iscijeli srce, rastani se, kreni!

Herman Hesse, iz *Igre staklenim perlama*

SUMMARY

Glycans constitute the most abundant and diverse form of the post-translational modifications. While genes unequivocally determine the structure of each polypeptide, there is no genetic template for the glycan part. Instead, hundreds of genes and their products interact in the very complex pathway of glycan biosynthesis which is further complicated by environmental influences. Therefore, the aim of this thesis was to determine the extent to which individual differences in immunoglobulin G and total plasma proteins glycosylation patterns reflect genetic versus environmental influences. A twin study design was used and study subjects were twins enrolled in the TwinsUK registry, a national register of adult twins. More than 4500 samples were analyzed by HILIC-UPLC (Hydrophilic Interaction Ultra Performance Liquid Chromatography). A high contribution of the genetic component to *N*-glycome composition was found. Variation in levels of 51 of the 76 IgG glycan traits studied was at least 50% heritable and only a small proportion of *N*-glycan traits had a low genetic contribution. Heritability of plasma *N*-glycome was also high, with half of the plasma glycan traits being at least 50% heritable. Further, epigenome-wide association (EWA) analysis showed that methylation levels at some genes are also implicated in glycome composition, both in those with high heritability and those with a lower genetic contribution. The study to investigate the potential role of glycosylation in kidney function was also conducted. Fourteen IgG glycan traits were associated with renal function in discovery population and remained significant after validation in an independent subset of monozygotic twins discordant for renal disease, reflecting difference in galactosylation, sialylation, and level of bisecting *N*-acetylglucosamine. Using the weighted correlation network analysis (WGCNA) for IgG glycan traits, a correlation between low back pain (LBP) and glycan modules was established. There was a weak positive correlation between pain phenotypes and "pro-antibody-dependent cell-mediated cytotoxicity (ADCC)" WGCNA glycan modules (high bisecting *N*-acetylglucosamine and low core fucose) and a weak negative correlation between pain phenotypes and "anti-ADCC" module (high core fucose, no bisecting *N*-acetylglucosamine). This suggests that glycans are promising candidates for biomarkers in many different diseases.

KEYWORDS: IgG glycome, total plasma glycome, *N*-glycosylation, glycan analysis, HILIC, twin study, heritability, chronic kidney disease, low back pain

SAŽETAK

Uvod: Glikozilacija je složen, visokospecifičan i strogo reguliran kotranslacijski proces kovalentnog vezanja složenih šećernih struktura na proteine i lipide. Dok geni nedvosmisleno određuju strukturu svakog polipeptida, za sintezu glikana ne postoji genski predložak. Umjesto toga, stotine gena i njihovih produkata sudjeluju u vrlo kompleksnoj biosintezi glikana koju okolišni utjecaji čine još složenijom. Stoga je cilj ovog doktorskog rada odrediti razmjer kojim genski i okolišni čimbenici utječu na *N*-glikane imunoglobulina G (IgG) i ukupnih glikoproteina plazme. Da bi se postigao navedeni cilj, upotrijebljena je studija na blizancima.

Ispitanici: Ispitanici su regrutirani *TwinsUK* registrom, najvećim registrom blizanaca u Velikoj Britaniji. Svi su ispitanici potpisali informirani pristanak o sudjelovanju u ispitivanju te za njega postoje dopuštenja odgovarajućih britanskih etičkih povjerenstava. Za ovo se istraživanje upotrijebila njihova krvna plazma – 90 μl po uzorku za analizu *N*-glikana IgG-a te 10 μl po uzorku za analizu *N*-glikana ukupnih proteina plazme.

Metode: IgG je izoliran iz krvne plazme uz pomoć protein G monolitne pločice. Glikani ukupnih proteina plazme pripremljeni su za analizu na jednak način kao glikani IgG-a. *N*-glikani su oslobođeni od proteina enzimom PNG-azom F te su obilježeni 2-aminobenzamidom (koji je fluorofor) u reakciji reduktivne aminacije. Fluorescentno obilježeni i pročišćeni *N*-glikani analizirani su HILIC-UPLC metodom (kromatografijom vrlo visoke djelotvornosti temeljenoj na hidrofilnim interakcijama, eng. *Hydrophilic Interaction Ultra Performance Liquid Chromatography*). Analizirano je više od 4500 uzoraka.
Epigenomska asocijacijska analiza provedena je pomoću Illumina 27k čipa.

Rezultati:

U prvoj studiji analizirani su glikani 220 monozigotnih (MZ) i 310 dizigotnih (DZ) blizanaca. Za analizu podataka upotrijebljen je klasični dizajn studije na blizancima kojim se uspoređuje varijacija unutar jedne grupe s varijacijom izvan te grupe (MZ i DZ blizanci). Ispitivao se utjecaj: aditivne genetike - što podrazumijeva sumu svih efekata jednog individualnog mjesta na kromosomu, zajedničnog okoliša i uvjeta života te individualnih odgovora na utjecaj okoliša, specifičnog za pojedinca. Velika genska komponenta

(heritabilnost, $h^2 \geq$ 50%) pokazana je za 51 od 76 glikanskih svojstava IgG-a. Nasuprot tome, samo je 12 glikanskih svojstava IgG-a pokazalo malu gensku komponentu ($h^2 < 35\%$). Heritabilnost plazmatskog *N*-glikoma također se pokazala velikom. Polovica plazmatskih glikanskih svojstava bila je barem 50% heritabilna. Prosječna heritabilnost IgG glikoma (53%, bez izvedenih svojstava) nešto je veća od prosječne heritabilnosti ukupnog plazmatskog glikoma (47%). Također je testirana povezanost između razine metilacije DNA i razine IgG glikana ($P<2 \times 10^{-6}$). Epigenomskom asocijacijskom analizom pronađeno je 5 značajnih asocijacija (koje odgovaraju dvama različitim genima) za 64 glikanska svojstava s $h^2 > 35\%$ i 5 značajnih asocijacija (koje također odgovaraju dvama različitim genima) za 12 glikanskih svojstava s $h^2 \leq 0.35$. Udio značajnih asocijacija bio je značajno veći ($P<0.005$) među glikanima s niskom heritabilnošću (42%), u usporedbi s glikanima s visokom heritabilnošću (6.2%).

Također je provedeno prvo istraživanje potencijalne uloge IgG glikozilacije u funkciji bubrega. Istražena je povezanost između IgG glikana i bubrežne funkcije kod 3274 ispitanika. Koristeći se linearnom regresijom korigiranom na kovarijable i višestruko testiranje, analizirana je korelacija između bubrežne funkcije iskazane kao procijenjena brzina glomerularne filtracije (eng. *estimated glomerular filtration rate,* eGFR) i 76 glikanskih svojstava IgG-a. Rezultati su replicirani na 31 paru MZ blizanaca diskordantnih za bubrežnu funkciju (blizanci svakog pojedinačnog para imaju različitu bubrežnu funkciju). Nakon toga, rezultati obje analize su metaanalizirani. U prvoj je analizi pronađena značajna povezanost ($P<6.5\times10^{-4}$) s bubrežnom funkcijom za 14 glikanskih svojstava, a ostala je značajna i nakon validacije. Ta glikanska svojstva pripadaju trima glavnim glikozilacijskim karakteristikama IgG-a: galaktozilaciji (pozitivno korelirana s eGFR-om), sijalinizaciji (pozitivno korelirana s eGFR-om) i razini račvajućeg *N*-acetilglukozamina (negativno korelirana s eGFR-om). Da bi se utvrdilo jesu li promjene u glikozilaciji kod kronične bubrežne bolesti ograničene na IgG ili su rezultat opće promjene glikozilacije više različitih proteina, istražena je povezanost između ukupnih plazmatskih *N*-glikana i eGFR-a u podskupini od 426 ispitanika. Nije pronađena razlika u plazmatskoj glikozilaciji, što upućuje na to da su prije pronađeni efekti izravna posljedica promijenjene glikozilacije IgG-a. Rezultati ove studije ukazuju na ulogu IgG-a u bubrežnoj funkciji i pružaju novi uvid u patofiziologiju kronične bubrežne bolesti te predlažu glikane kao moguće biomarkere za tu bolest.

Bol u donjem dijelu leđa (eng. *low back pain*, LBP) učestali je iscrpljujući poremećaj kojemu su etiologija i patogeneza slabo razjašnjeni. U sklopu ovog doktorskg rada provedena je prva analiza povezanosti između LBP-a i plazmatskih IgG *N*-glikana u skupini od 4511

blizanaca koji su ispitani za LBP i oštećenje intervertebralnog diska, kao mogućeg uzroka LBP-a. Koristeći se WGCNA (eng. *weighted correlation network analysis*) metodologijom, provedena je mrežna anliza razina IgG glikana kod blizanaca, da bi se uspostavili klasteri koreliranih glikana. Identificirano je sedam modula koreliranih glikana koji odražavaju funkcionalno povezane grupe glikana i pokazuju poklapajuće biološke aktivnosti. Pronađene su povezanosti između tih klastera i fenotipova boli kod blizanaca s LBP-om. Opažena je pozitivna korelacija između fenotipova boli i „pro-stanična citotoksičnost ovisna o protutijelima (eng. *antibody-dependent cell-mediated cytotoxicity*, ADCC)“ WGCNA glikanskog modula (visoka razina račvajućeg *N*-acetilglukozamina i niska razina sržne fukoze) te negativna korelacija između fenotipova boli i „anti-ADCC“ modula (visoka razina sržne fukoze, bez račvajućeg *N*-acetilglukozamina). Razine četiriju glikanskih svojstava, koja su zastupljena u ta dva WGCNA modula, statistički su se značajno razlikovale kod MZ blizanaca diskordantnih za LBP. Također, pokazan je trend prema većoj prevalenciji sistemskih upalnih poremećaja kod blizanaca s niskom razinom fukoziliranih glikana i visokom razinom glikana s račvajućim *N*-acetilglukozaminom. Sržna fukozilacija je „sigurnosni prekidač“ koji sprječava prekomjerni ADCC, stoga ovi rezultati upućuju na sudjelovanje ADCC-a i posljedične upale u patogenezi LBP-a. Nije pronađena korelacija između oštećenja intervertebralnih diskova i razine glikana, što upućuje na to da upala možda ne sudjeluje u oštećenju diskova. Ovi rezultati pružaju novi uvid u razumijevanje složene patofiziologije LBP-a i ističu glikane kao moguće biomarkere za neke podvrste LBP-a, u patogenezu kojih je uključena upala.

Zaključak: Unatoč nedostatku strogoga genskog predloška, varijabilnost razine glikana ponajprije je posljedica genske pozadine i specifičnih patofizioloških procesa pa bi glikani mogli postati obećavajući biomarkeri za mnoge bolesti. Metodom upotrijebljenom za analizu glikana u ovome radu još uvijek se rutinski ne koristi u kliničkoj praksi. Međutim, u skorijoj budućnosti, napredovanjem tehnologije koja omogućuje visokoprotočnu analizu glikanskih profila, glikani bi mogli postati atraktivni i klinički isplativi biomarkeri. Jednako tako, širenje znanja o glikozilaciji proteina moglo bi potaknuti nove uvide u patofiziologiju bolesti i razvoj novih terapeutika.

KLJUČNE RIJEČI: IgG glikom, ukupni plazmatski glikom, *N*-glikozilacija, analiza glikana, HILIC, studija na blizancima, kronična bubrežna bolest, bol u donjem dijelu leđa

TABLE OF CONTENTS

1. INTRODUCTION 1

1.1. The importance of glycosylation 2

1.2. Human plasma protein *N*-glycosylation 4

1.3. Immunoglobulin G *N*-glycosylation 5

1.3.1. Function and structure of immunoglobulin G 5

1.3.2. Interaction between Fc glycans and FcγRs is critical for IgG function 6

1.3.3. IgG N-glycosylation changes with age and in many diseases 8

1.4. High-throughput glycomics 8

1.5. Twin studies 10

1.5.1. Disadvantages of twin studies 11

2. Glycosylation of Immunoglobulin G: Role of Genetic and Epigenetic Influences 13

3. Glycosylation Profile of IgG in Moderate Kidney Dysfunction 22

4. The Association Between Low Back Pain and Composition of IgG Glycome 32

5. GENERAL DISCUSSION 44

6. CONCLUSIONS 50

7. REFERENCE LIST 52

8. SUPPLEMENTAL DATA 64

8.1. Supplemental data 1: Glycosylation of Immunoglobulin G: Role of Genetic and Epigenetic Influences 65

8.2. Supplemental data 2: Table 1 71

8.3. Supplemental data 3: Glycosylation Profile of IgG in Moderate Kidney Dysfunction 73

8.4. Supplemental data 4: The Association Between Low Back Pain and Composition of IgG Glycome 79

9. BIOGRAPHY 90

1. INTRODUCTION

1.1. **The importance of glycosylation**

Glycosylation is a complex, highly specific and strictly regulated cotranslational process of covalent bonding of complex sugar structures on proteins and lipids [1,2]. Changes of glycosylation can strongly effect the structure and function of proteins. Glycans have numerous important roles such as modulation of protein degradation, folding and secretion, cell signalling, immune function and transcription [3–5]. Multicellular life without glycans is not possible and the complete absence of glycans is embryologically lethal [6]. Most proteins in nature are glycoproteins [7]. We can find them in plants, animals and microorganisms. More than half of all human proteins and nearly all membrane and extracellular proteins are glycosylated [8]. Glycoproteins can be found in soluble form and in membranes, in all compartments of the cell and also in extracellular space. Different glycans can be attached to multiple glycosylation sites on a single protein, thus each glycoprotein comes as a mixture of numerous glycoforms. Monosaccharides that form glycans are linked via glycosidic bonds, which can be α linkages or β linkages, depending on the relationship of the oxygen from hydroxyl group to the anomeric carbon. These linkages give a different structural properties and biological functions to the sequences that are otherwise identical in composition (e.g. cellulose and starch) [9].

Due to the nature of linkage, by which they are attached to polypeptide backbones, glycans in eukaryotes are divided in *N*-linked, *O*-linked and *C*-linked glycans [10]. A *N*-linked glycan is an oligosaccharide covalently linked to an asparagine residue of a polypeptide chain which generally occurs at the sequon: Asn-X-Ser/Thr. These glycans are transferred to protein moiety on the luminal side of the endoplasmatic reticulum (ER) membrane. *N*-glycans have a common core region made of five monosaccharides and they further differ by the subsequent sequence on the basis of which they form three subgroups: high mannose - only mannose residues are attached to the core, complex - two or more antennae are attached to core via *N*-acetylglucosamine (GlcNac) and hybrid type *N*-linked glycans - mannose residues attached to the Manα1,6 arm and one or two antennae to the Manα1,3 arm (Figure 1). An *O*-linked glycan is usually linked to the polypeptide chain via *N*-acetylgalactosamine (GalNAc) to a serine or threonine residue and can exist as variety of different structural classes [10]. *C*-linked glycans are a rare form of glycosylation which involves the addition of an α-mannopyranose to the C2-carbon of the indole ring of tryptophan [11].

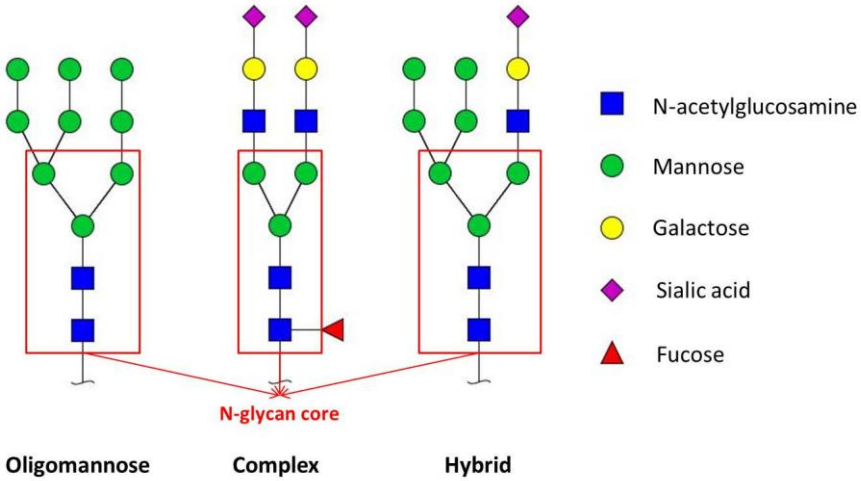


Figure 1. Examples of three *N*-glycan types: oligomannose (only mannose residues are attached to the core), complex (two or more antennae are attached to core via *N*-acetylglucosamine) and hybrid glycans (a mixture of the two previous types - it has one or more oligomannose branches and one branch of complex structure). All *N*-glycans share a common core (marked with red rectangle). From: Gornik, O., Keser, T. & Lauc, G. in *Sample Prep. Tech. Soil, Plant, Anim. Samples* (ed. Mičić, M.) (Humana Press, 2016).

While genes unequivocally determine the structure of each polypeptide, there is no direct genetic template for the glycan part [12]. Instead, glycans are shaped by complex dynamic interactions between hundreds of enzymes, transcription factors, ion channels and other proteins. That is further complicated by both direct environmental influence (nutrition, hormonal status, etc.) and epigenetic memory of past environmental effects (altered gene expression) [13–16]. Glycosylation is known to be affected by factors such as: type of glyco-enzymes, their localization and expression levels, abundance and trafficking of glycoprotein substrates and activated sugar donors concentrations [17,18].

Probably the most important feature of glycoproteins is their heterogeneity. It can be manifested from minor to considerable differences through higher branching, loss of monosaccharides from one of glycan branches, through absence or presence of certain monosaccharide such as sialic acid, fucose, *N*-acetylglucosamine or through type of linkage between sugars [19]. Variability in terminal glycan antennae is common and recent studies demonstrated significant variation in glycome composition between individuals [20–22]. Many diseases are associated with changes in glycan structures, e.g. congenital disorders of glycosylation, cancer, autoimmune diseases (rheumatoid arthritis, systemic lupus

erythematosus, inflammatory bowel disease) and AIDS [23,24]. Rapid glycosylation changes, associated with occurrence, severity and outcome of pathological processes, have a potential for good biochemical markers and as novel therapeutic targets [19,25].

1.2. **Human plasma protein *N*-glycosylation**

Most human plasma proteins, except for albumin and CRP (C-reactive protein), are modified by glycans [26]. Knežević et al. [21] performed the first large-scale analysis of human plasma glycome, revealing a high variability in a glycome composition between individuals with the median difference between the minimal and maximal values of glycans being over six-fold. At the same time, individual plasma glycome appears to change very little in a healthy individual, even after a prolonged period of time, which indicates stable long-term regulation of the glycosylation machinery [27]. It is only when the homeostasis of a person changes, by lifestyle or pathological conditions, that the glycosylation will change notably [28].

Variations observed in a human glycome are probably a combination of genetic differences and environmental factors. Several studies have already shown that there are consistent genetic factors that affect circulating levels of some *N*-glycans [29–31]. Heritability is one of the most basic and often one of the first analyses to be made in a genetic study, since it represents the proportion of the trait variance that can be attributed to genetic factors [32]. A broad range of variation in heritability levels of plasma glycans has been shown in one study, from insignificant or very low to over 50% for some glycans [21]. Pedigree information was used to calculate this heritability estimates. Their interpretation is usually complicated by differences in methodological approaches and sampling schemes and variation in trait values over time. Population-specific differences can all influence heritability values, resulting in a wide range of heritability estimates in different studies [32]. The heritability of glycome has never been estimated with twin studies. The advantage of twin studies is that the total variance can be split up into genetic, shared or common environmental, and unique environmental components, enabling an accurate estimation of heritability [33].

The fact that there are hundreds of genes involved in the complex glycan metabolic pathways argues in favor of a strong genetic influence, but environmental effects on glycan structures have also been reported [34–37]. Large studies which involve thousands of individuals have identified that glycosylation correlates with age, sex and lifestyle [21,28,30]. However,

besides age, which significantly affected galactosylation, all other environmental factors individually accounted only for a small fraction of the observed variance, thus the main source of glycome variation between individuals is still not known.

Since plasma proteins originate from different tissues and organs, their properties are affected by the physiological or pathological conditions of these tissues and organs, indicating that plasma proteins and their glycans could be good biomarkers for monitoring various health conditions [26]. Plasma is an easily obtainable biofluid and hence its glycosylation analysis could be of considerable interest for using in clinics. Although, when analyzing total plasma *N*-glycome at the released glycan level, it is not directly apparent whether an observed change originates from a change in relative protein abundance, in the relative glycoforms of a specific protein, or whether it reflects a general regulatory effect influencing the glycosylation of many different glycoproteins [38]. Therefore, glycosylation analysis of individual proteins of human plasma could help in understanding the total plasma *N*-glycomic changes.

1.3. Immunoglobulin G *N*-glycosylation

1.3.1. Function and structure of immunoglobulin G

Immunoglobulins (Igs) have a significant role in the adaptive immune system by defending the body against invading pathogens. All five classes of human Igs (IgG, IgM, IgA, IgD and IgE) are glycoproteins which are produced by B cells[39,40]. IgG is the most abundant Ig class in the human blood (approx. 10 mg/ml, 15–20% of serum glycoproteins) and a major effector molecule of the humoral immune response. There are four subclasses of human IgG: IgG1, IgG2, IgG3 and IgG4. They are all glycoproteins composed of two heavy and two light chains linked together by interchain disulphide bonds. The two light chains together with the parts of the heavy chains (VH and CH1 domains) form two Fab moieties which are linked by a flexible hinge region to one Fc moiety formed by the remainders of the two heavy chains (CH2 and CH3 domains) [41,42]. The effectiveness of IgG antibodies comes from two functional properties: first, the ability of the Fab regions to specifically recognize and bind the target antigen; and second, the ability to induce different immune system effector mechanisms through interaction of the Fc region with: Fc gamma receptors (FcγRs), the C1q component of complement and the neonatal receptor (FcRn) [43].

1.3.2. Interaction between Fc glycans and FcγRs is critical for IgG function

FcγRs play a crucial role in induction of IgG effector mechanisms. Human FcγRs are divided into the three main groups: I, II and III. FcγRII and FcγRIII are further divided in subgroups: IIa, IIb, IIc and IIIa, IIIb. FcγRs differ in their affinity for the Ig Fc-part. The FcγRI is usually described as a high affinity receptor, whereas the FcγRII and the FcγRIII are referred to as receptors with low and moderate affinity respectively. In view of initiated immune response, FcγRs can be further classified as activating or inhibitory. FcγRIIb is the only inhibitory receptor and the rest are activating receptors. Binding of an IgG or immune complexes to activating FcγRs induce: antibody dependent cell-mediated cytotoxicity (ADCC), phagocytosis and endocytosis and it can promote antigen presentation and release of pro-inflammatory mediators. FcγRIIb modulates the immune response by inhibiting the activation of activating receptors. Thus, immune responses are balanced between activating and inhibitory functions of the FcγRs [43].

Each heavy chain of IgG carries a single covalently attached *N*-glycan at the highly conserved asparagine 297 residue in each of the CH2 domains of the Fc region of the molecule [44]. The Fc *N*-glycans are biantennary complex-type structures which are mostly core-fucosylated and may contain a bisecting GlcNAc and a small portion of sialic acid. The majority of IgG *N*-glycans are attached to the heavy chains of the Fc region, but about 20% of polyclonal human IgG molecules also contain *N*-glycans within the Fab regions of the light chain, the heavy chain or both [45]. Fab glycosylation is restricted to the variable domains and it was shown to affect stability, half-life, activity and binding characteristics of IgG, however its function is still poorly understood [46].

Fc glycans are essential structural components of the IgG molecule and even minor changes in glycan composition can have a profound influence on IgG effector functions by modulating binding to Fc receptors [47] (Figure 2). Many important functional effects of alternative IgG glycosylation have been described [48]. Glycans that lack terminal galactose activate complement and make IgG pro-inflammatory, while the addition of galactose decreases inflammatory potential of IgG [49–51]. Further extension of IgG glycans by the addition of sialic acid dramatically changes the physiological role of IgG, converting it from a pro-inflammatory into an anti-inflammatory agent. Terminal α2,6-sialylation of IgG glycans decreases the ability of IgG to bind to activating FcγRs and promotes recognition by the C-type lectin dendritic cell-specific intercellular adhesion molecule (ICAM)-grabbing

non-integrin (DC-SIGN), which increases expression of inhibitory FcγRIIB and is anti-inflammatory [52,53]. However, these findings have not been confirmed in all studies [54–56].

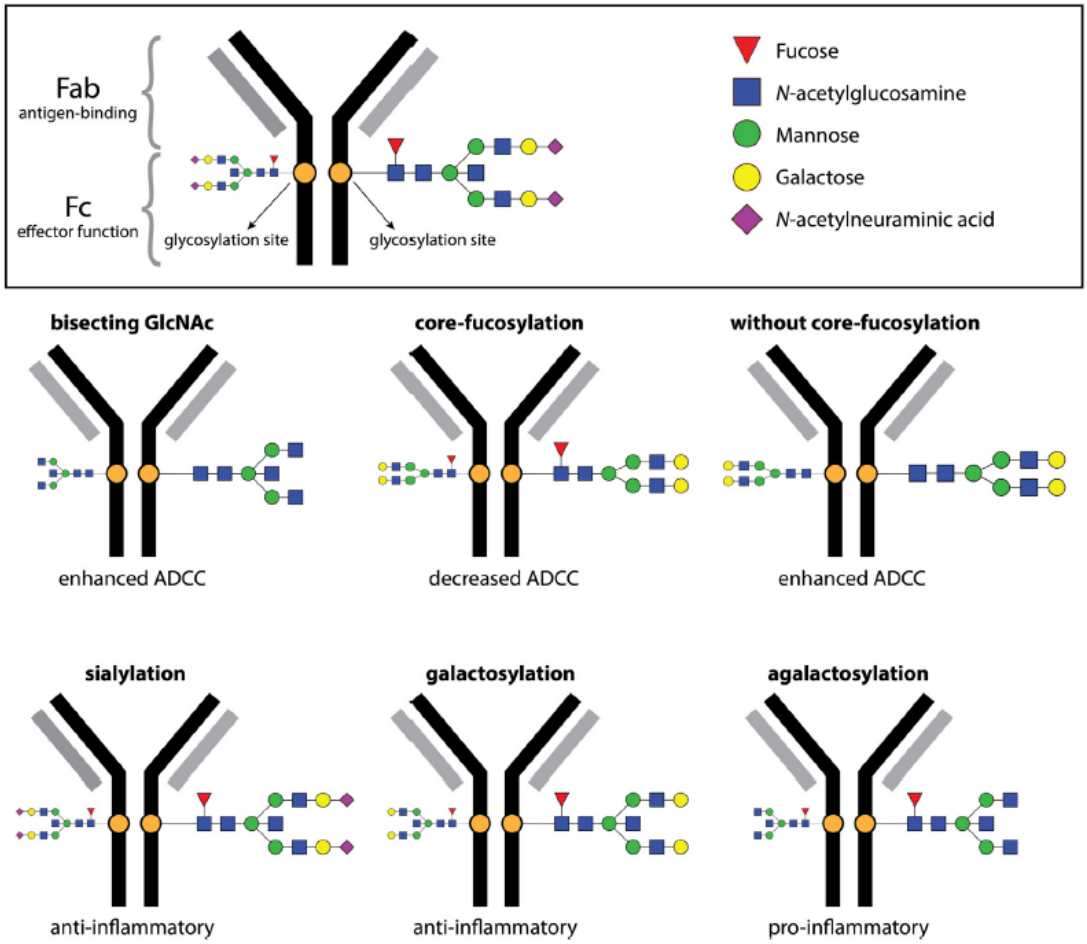


Figure 2. Changes in Fc glycan composition influence IgG effector functions by modulating binding to Fc receptors. From: Lauc, G., Pezer, M., Rudan, I. & Campbell, H. Mechanisms of disease: The human N-glycome. *Biochim. Biophys. Acta - Gen. Subj.* (2015).

Another example of Fc glycans influencing the IgG function is the role of core fucose in the modulation of ADCC: IgG containing glycans that lack core fucose have over 50-fold increased affinity for FcγRIIIa and FcγRIIIb and are therefore much more efficient in activating ADCC than fucosylated glycoforms of the same molecule. By binding to this activating Fc receptor expressed primarily on natural killer (NK) cells antibodies initiate ADCC which leads to destruction of target cells [57]. Subsequently, improved ADCC by increased interaction with Fc receptors was shown in CHO cells transfected with the human β1,4-Nacetylglucosaminyltransferase III (GnT-III) gene which adds bisecting GlcNAc [58,59]. The addition of bisecting GlcNAc, a relatively early event in glycoprotein processing, inhibits α1,6-fucosyltransferase (FUT8) and the addition of core fucose [60]. However, it seems that the

lack of core fucose and not the presence of bisecting GlcNAc, has the most critical role in enhanced ADCC [61]. Due to drastic enhancement of ADCC, afucosylated monoclonal antibodies have a strong therapeutic potential in anti-cancer therapy [62].

In addition to functional aspects, Fc glycosylation has also been shown to play an important role in maintaining the structural integrity, stability and solubility of IgGs [63,64].

1.3.3. IgG *N*-glycosylation changes with age and in many diseases

Several IgG glycans change considerably with age and the combination of only three glycans can explain up to 58% of variance in chronological age, significantly more than other markers of biological age like telomere lengths [65]. The most influenced feature of IgG glycosylation is galactosylation, which decreases to less than 50% of its maximal value through lifetime. The decrease in IgG galactosylation with age was initially reported almost 30 years ago [66] and was replicated in a number of subsequent studies [65,67].

It is well known that IgG glycosylation pattern is changing in many diseases. Similar to healthy individuals, IgG heterogeneity in diseased individuals is also most commonly expressed in galactosylation levels. More than 30 years ago, low IgG galactosylation was associated with rheumatoid arthritis and osteoarthritis [68] and since then many different studies have reported reduction of galactosylation in a number of different pathological conditions: infectious diseases (hepatitis C infection [69], HIV infection [70]), autoimmune diseases (rheumatoid arthritis [68,71–73], juvenile chronic arthritis [74,75], Chron's disease [24,76], ulcerative colitis [24,76], systemic lupus erythematosus [77], myositis [78]) and cancer (prostate cancer [79], lung cancer [80], ovarian cancer [81], gastric cancer [82], breast cancer [83]). The exact meaning of these changes is still difficult to interpret and further studies are necessary to clarify variation and regulation of IgG glycosylation in an immune response.

1.4. High-throughput glycomics

The existence of many thousands of different glycans attached to human proteins, multiple glycosylation sites, large variety of the attached glycans in a single glycosylation site and structural complexity of glycans (differences in monosaccharide composition, anomeric state, linkage of the subunits, branching and linkage to the peptide part of a glycoprotein), all contribute to complexity of glycan structural analysis [23,84,85]. Therefore, the knowledge about

the role of glycans in disease mechanisms is lagging significantly behind the knowledge about the role of genes and proteins. However, recently a major progress has been made and several high-throughput analytical techniques for glycan analysis have been developed [22,86–91].

Glycan analysis may involve: whole glycoprotein analysis, analysis of glycopeptides obtained after enzymatic treatment of the glycoprotein, analysis of released glycans obtained after chemical or enzymatic treatment of the glycoprotein and monosaccharide analysis [11]. Analysis of released glycans is the most widely used as it provides a convenient way to obtain information on the various populations of glycans present on the protein. In general, there are three major analytical techniques that are used for glycan analysis: liquid chromatography (high performance – HPLC and ultra performance - UPLC), mass spectrometry (MS) and capillary electrophoresis (CE) [92]. In most cases UPLC and CE enable reliable relative glycan quantification, but separation of glycans is rarely complete and each peak is usually a mixture of signal contributions originating from different glycans [93]. MS analysis can provide more structural details, but is generally less quantitative. An additional problem of MS lies in the fact that signal response factors are different for different glycans [92].

One of the major bottlenecks in large scale proteomics and glycomics studies is protein purification from a great number of samples. The most widely used technique for that purpose is affinity chromatography using immobilized affinity ligand. Protein G is frequently used as bioaffinity ligand for the IgG purification from plasma or serum [94,95]. All four subclasses of human IgG strongly bind to protein G through their Fc fragments [96].

To study the glycosylation of the protein, its structure or function, it is often needed to remove its glycans. The most commonly used enzyme for deglycosylation of glycoproteins is peptide-*N*-glycosidase F (PNGase F) [21]. This enzyme efficiently releases all asparagine-linked glycans unless they are α1,3-core fucosylated - a modification usually found in plants and insects [97].

There is no natural chromophore in the carbohydrate molecules. Therefore, a wide range of techniques have been developed for the detection of glycans. The released glycans are most commonly fluorescently tagged (with fluorophores such as: 2-aminobenzoic acid, 2-aminobenzamide, 2-aminopyridine, 1-aminopyrene-3,6,8-trisulfonic acid) at their reducing end by reductive amination [98–100] and then separated by LC or CE, followed by fluorescence

detection or mass spectrometry. They can also be derivatized to improve the ionization efficiency and minimize in-source or post-source decay in MS analysis [91,101].

Liquid chromatography is broadly used for analysis of both neutral and charged glycans in a single separation [21,22,81,86,93]. Glycans are predominantly separated using hydrophilic interaction liquid chromatography (HILIC) mode, however, high-pH anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD), weak anion exchange (WAX), porous graphitic carbon (PGC) and reverse phase (RP) chromatographies can also be employed [11]. In HILIC, glycans are resolved based on differences in the hydrophilicity and hydrophobicity. Glycan size, charge, composition, linkage and arm specificity are all contributing to retention times [86,102]. HILIC enables the separation of structural isomers, which is one of its main advantages [93].

None of the currently available analytical techniques are capable of performing a detailed structural analysis of protein glycosylation in a single step. However, a multistep process which combines existing methods can provide detailed characterization of protein glycan profile in complex biological samples.

1.5. Twin studies

Twin studies are often used for the estimation of heritability. They are based on analysis and comparison of a given trait in monozygotic (MZ) and dizygotic (DZ) twins, to determine the extent to which genetic and environmental factors influence the given trait. MZ twins share all of their genetic background and they grow up sharing many factors in their environment. However, even MZ twins growing up together do not share all possible environmental experiences - some experiences are unique to each twin. The classical twin-design contains only twins raised in their biological family. If MZ twins are considerably more similar than DZ twins (which is found for most traits), this implicates that genes play an important role in a trait. The similarity between dizygotic twins indicates that shared environmental factors have an important influence on a trait [103,104].

Heritability is estimated using structural equation modelling to break down the observed phenotypic variance into three sources of variation: additive genetic variance (A), shared/common environmental variance (C), and non-shared/unique environmental variance (E). Additive genetic influences are indicated when MZ twins are more similar than DZ

twins. The common environmental component estimates the contribution of family environment which is assumed to be equal in both MZ and DZ twin pairs, whereas the unique environmental component does not contribute to twin similarity, rather it estimates the effects that apply only to each individual and includes measurement error. The equal environment assumption across zygosities implies that any greater similarity between MZ twins than DZ twins is attributed to greater sharing of genetic influences. The best fitting model (among ACE, AE, CE, and E models) is used to estimate heritability, defined as the proportion of the phenotypic variation attributable to genetic factors [103,105].

Twins are also useful in case–control studies because MZ twins offer the possibility of carrying out the ideal case–control study, as they are perfectly matched for genotype and family background [104].

1.5.1. Disadvantages of twin studies

Twin studies usually presume that genes and environment have independent and distinct contributions to a trait. Nevertheless, interactions between genes and environment could influence the development and intensity of traits. With the use of twins it is not possible to examine the effects of both shared environment and gene-environment interaction at the same time. This problem can be solved by including additional siblings in the study design.

Gene-environment interactions can be mediated via methylation. DNA methylation is one of the main epigenetic mechanisms, which attribute to chemical instructions for gene activity that do not alter DNA sequences [106]. Differences in methylation, which can be an outcome of unique environmental influences, may create phenotypic differences between MZ twins. Therefore, ignoring this issue could lead to a false conclusion of direct environmental causes to explain phenotypic discordance in MZ twins [107]. However, availability of new and relatively inexpensive technology to determine changes in gene expression can easily solve this problem. Another solution would be to study relatively young MZ twin pairs, because epigenetic differences between MZ twins seem to become apparent later in life [108].

Also, one thing to consider when using twin studies is that their results cannot be directly generalized to the general population. Twins are not a random sample of the general population. They differ in their developmental environment and parents' genetic factors may lead to a higher incidence of twinning. Therefore, twin study results should be validated by

demonstrating that MZ twin samples are comparable to population-based samples of singletons [109].

However, when all the precautions are taken, twin studies provide an ideal model to estimate the proportion of variance in a trait attributable to genetic variation, versus the proportion that is due to shared environment or unshared environment.

2. Glycosylation of Immunoglobulin G: Role of Genetic and Epigenetic Influences

Glycosylation of Immunoglobulin G: Role of Genetic and Epigenetic Influences

Cristina Menni[1*☯], Toma Keser[2☯], Massimo Mangino[1], Jordana T. Bell[1], Idil Erte[1], Irena Akmačić[3], Frano Vučković[3], Maja Pučić Baković[3], Olga Gornik[2], Mark I. McCarthy[4,5], Vlatka Zoldoš[6¶], Tim D. Spector[1¶], Gordan Lauc[2,3¶], Ana M. Valdes[1,7¶]

1 Department of Twins Research and Genetic Epidemiology, Kings College London, London, United Kingdom, **2** Faculty of Pharmacy and Biochemistry, University of Zagreb, Zagreb, Croatia, **3** Glycobiology Laboratory, Genos, Zagreb, Croatia, **4** Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Oxford, United Kingdom, **5** Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, **6** Faculty of Science, University of Zagreb, Zagreb, Croatia, **7** Academic Rheumatology, University of Nottingham, Nottingham, United Kingdom

Abstract

Objective: To determine the extent to which genetic and epigenetic factors contribute to variations in glycosylation of immunoglobulin G (IgG) in humans.

Methods: 76 *N*-glycan traits in circulating IgG were analyzed by UPLC in 220 monozygotic and 310 dizygotic twin pairs from TwinsUK. A classical twin study design was used to derive the additive genetic, common and unique environmental components defining the variance in these traits. Epigenome-wide association analysis was performed using the Illumina 27k chip.

Results: 51 of the 76 glycan traits studied have an additive genetic component (heritability, h^2)≥ 0.5. In contrast, 12 glycan traits had a low genetic contribution (h^2<0.35). We then tested for association between methylation levels and glycan levels (P<2 x10^{-6}). Among glycan traits with low heritability probe cg08392591 maps to a CpG island 5' from the *ANKRD11* gene, a p53 activator on chromosome 16. Probe cg26991199 maps to the *SRSF10* gene involved in regulation of RNA splicing and particularly in regulation of splicing of mRNA precursors upon heat shock. Among those with high heritability we found cg13782134 (mapping to the *NRN1L* gene) and cg16029957 mapping near the *QPCT* gene to be array-wide significant. The proportion of array-wide epigenetic associations was significantly larger (P<0.005) among glycans with low heritability (42%) than in those with high heritability (6.2%).

Conclusions: Glycome analyses might provide a useful integration of genetic and non-genetic factors to further our understanding of the role of glycosylation in both normal physiology and disease.

Citation: Menni C, Keser T, Mangino M, Bell JT, Erte I, et al. (2013) Glycosylation of Immunoglobulin G: Role of Genetic and Epigenetic Influences. PLoS ONE 8(12): e82558. doi:10.1371/journal.pone.0082558

Editor: Gunnar F Kaufmann, The Scripps Research Institute and Sorrento Therapeutics, Inc., United States of America

Received July 23, 2013; **Accepted** October 23, 2013; **Published** December 6, 2013

Funding: TwinsUK was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013). The study also receives support from the National Institute for Health Research (NIHR) Clinical Research Facility at Guy's & St Thomas' NHS Foundation Trust and NIHR Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. TS is an NIHR senior Investigator and is holder of an ERC Advanced Principal Investigator award. SNP Genotyping was performed by The Wellcome Trust Sanger Institute and National Eye Institute via NIH/CIDR. This work was also supported by the European Commission HighGlycan (contract #278535), IBD-BIOM (contract#305479), Integra-Life (contract #315997) and MIMOmics (contract #305280) grants. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have read the journal's policy and have the following conflicts: GL declares that he is a founder and owner, and IA, FV, MPB declare that they are employees, of Genos Ltd, which offers commercial service of glycomic analysis and has several patents in this field. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: cristina.menni@kcl.ac.uk

☯ These authors contributed equally to this work.

¶ These authors also contributed equally to this work.

Introduction

Glycans constitute the most abundant and diverse form of the post-translational modifications. All cell surface and secreted glycoproteins that contain appropriate sequences (Asn-X-Ser/Thr where X is any amino acid except proline) can potentially acquire N-linked oligosaccharides (N-glycans) while they travel through the endoplasmic reticulum and the Golgi compartments [1]. Glycans can influence disease development in many syndromes such as congenital disorders of

Table 1. Demographic characteristics of the study population, *mean (SD)*.

Phenotype	MZ	DZ	P
N	440	610	
age, yrs	58.71(9.37)	57.83(9.61)	0.14
BMI, kg/m^2	26.65(4.85)	26.50(4.68)	0.60

doi: 10.1371/journal.pone.0082558.t001

glycosylation, cancer, rheumatoid arthritis and AIDS [2]. Glycans are crucial for the immune system, as some of the most important interactions between the immune system and viruses and bacteria are mediated by protein-glycan interactions. Moreover, glycans are key in the recognition of non-self events and an altered glycome can lead to autoimmune disorders [3]. The biological functions of glycans go from basic structural roles to development, protein folding and immune response. Glycosylation is known to be affected by factors such as sugar nucleotide concentration, type of glyco-enzymes and their expression levels [1].

While genes unequivocally determine the structure of each polypeptide, there is no genetic template for the glycan part [4]. Instead, hundreds of genes and their products interact in the complex pathway of glycan biosynthesis resulting in a very complex biosynthetic pathway that is further complicated by both direct environmental influence (nutrition, hormonal status, etc) and epigenetic memory of past environmental effects (altered gene expression) [5-8].

Some recent studies [9,10] demonstrated large variability of plasma proteins and immunoglobulin G (IgG) *N*-glycome composition in a population. However, longitudinal studies in individuals also revealed very high temporal stability of the individual plasma glycome [11] indicating stable long-term regulation of the glycosylation machinery.

Several studies have already shown that there are consistent genetic factors that affect circulating levels of some IgG *N*-glycans [12-14]. However, the extent to which genetic and non-genetic factors affect glycan levels is still unexplored.

In this study we aim to determine the extent to which individual differences in IgG glycosylation pattern reflect genetic *versus* environmental influences by estimating heritability in a cohort of female twins. We, then, hypothesized that IgG glycosylation pattern with low heritability ($h^2 \leq 0.35$) might be epigenetically mediated and we explored the relationship with epigenetic data.

Results

The study cohort comprised 220 monozygotic (MZ) and 310 dizygotic (DZ) twin pairs, and the baseline characteristics are presented in Table **1**. MZ and DZ twin pairs were not significantly different for age and body mass index (BMI), and all of them were females. The number of twin pairs used in this study is sufficient to detect with 95% power heritability of 0.4 or higher with $P<0.05$ for a range of shared environmental contributions from low (0.1) to high (0.5) [15].

In total 76 glycan traits were studied which were derived from 24 directly measured glycan structures and derived measures. A detailed description of the glycan structures and the derived measures can be found in Lauc et al. [14]. For example, the Fc region of IgG contains two highly conserved asparagine (Asn-297) residues where complex core-fucosylated glycans such as FA2 (GP4) , FA2G1 (GP8, GP9) and FA2G2 (GP14) are found [14]. These carbohydrates are critical components of the Fc-Fcγ receptor interaction [1].

The heritability analyses for the 76 IgG patterns are presented in Table **2**. The best fitting model for the majority of the IgG pattern was the AE model (including additive genetic and non-shared environment), ascribing the total variance to additive genetic factors and non-shared environmental factors, with heritability estimates ranging from 0.49 for GP21 (which reflects the proportion of the A2G2S2 glycan, i.e. disialylated, digalactosylated, bi-antennary N-linked glycans; see Table **S1** for description of glycan codes) to 0.80 for GP8n (which reflects the proportion of the monogalactosylated isomer FA2[6]G1 relative to all the neutral glycans). The ACE model (including additive genetic, common environment and unique environments) was the best fitting model for 18 glycan traits with heritability estimates ranging from 0.19 for FtotalS1/ FtotalS2 to 0.52 for FBG0n/G0n. In contrast, 7 glycan traits showed no clear genetic influence and appeared to be affected primarily by common and unique environmental factors. We have classified these 7, added to the 5 glycan traits in Table **2** with an additive component under 0.35, as “low heritability”.

51 of the 76 glycan traits (67%) studied have an additive genetic component of 0.5 or greater, meaning that at least half of the variance in their levels is determined by genetic factors, 5 glycan traits had a genetic contribution under 0.35. Heritability estimates for the remaining glycan traits varied between 0.35 and 0.5.

We hypothesized that IgG glycans showing none or low genetic influences could be epigenetically mediated and we compared their levels with genome-wide DNA methylation profiles from the Illumina HumanMethylation27 DNA Analysis BeadChip assay in 127 MZ and DZ female twins [16]. The analyses were adjusted for age, sex, BMI, methylation chip, sample position on methylation chip, and family relatedness.

Among glycan traits with a low heritability we identified 2 CpG-sites (cg08392591, cg26991199) at which DNA methylation levels were associated with levels of 4 IgG glycan traits with $P < 2 \times 10^{-6}$ (the Bonferroni cut off for array-wide significance; see Table **3**). Probe cg08392591 maps to a CpG island 5’ from the *ANKRD11* gene on chromosome 16 while cg26991199 maps to *SRSF10* gene on chromosome 1. *ANKRD11* is a p53 activator, while *SFRS10* is involved in constitutive and regulated RNA splicing and in particular is involved in regulation of splicing of mRNA precursors upon heat shock.

Among the 64 glycan traits that had heritabilities above 0.35 we observed five array-wide significant hits (Table **3**) for the two probes, cg13782134 (mapping to the *NRN1L* gene) and cg16029957 (mapping near the *QPCT* gene). *NRN1L* encodes a neuritin-like protein precursor and its role in immunoglobulin glycosylation is unclear. *QPCT* encodes the human pituitary

Table 2. Heritability estimates and 95% confidence intervals for IgG glycan traits adjusted for age and batch.

	MZ		DZ					
Glycan Trait	**Mean(SD)***	**ICC[95%CI]**	**Mean(SD)***	**ICC[95%CI]**	**Best model**	**A[95%CI]**	**C[95%CI]**	**E[95%CI]**
GP1	0.1(0.06)	0.54[0.44,0.63]	0.11(0.06)	0.38[0.28,0.47]	AE	0.58[0.49,0.65]		0.42[0.35,0.51]
GP2	0.48(0.27)	0.68[0.61,0.75]	0.53(0.34)	0.24[0.13,0.34]	AE	0.72[0.64,0.48]		0.28[0.22,0.36]
GP4	19.41(6.45)	0.73[0.66,0.79]	19.26(5.63)	0.39[0.29,0.49]	AE	0.70[0.64,0.75]		0.30[0.25,0.36]
GP5	0.32(0.15)	0.63[0.54,0.71]	0.4(0.17)	0.41[0.32,0.50]	AE	0.72[0.65,0.77]		0.28[0.23,0.35]
GP6	5.33(1.77)	0.76[0.70,0.82]	5.32(1.68)	0.33[0.23,0.43]	AE	0.75[0.69,0.79]		0.25[0.21,0.31]
GP7	0.54(0.24)	0.66[0.58,0.73]	0.62(0.31)	0.33[0.23,0.43]	AE	0.73[0.67,0.79]		0.27[0.21,0.33]
GP8	18.97(1.74)	0.73[0.67,0.80]	19.19(1.83)	0.30[0.20,0.41]	AE	0.74[0.68,0.79]		0.26[0.21,0.32]
GP9	9.98(1.42)	0.74[0.68,0.80]	9.99(1.44)	0.41[0.32,0.51]	AE	0.75[0.69,0.79]		0.25[0.21,0.31]
GP10	6(1.21)	0.76[0.71,0.82]	5.99(1.19)	0.37[0.28,0.47]	AE	0.76[0.70,0.80]		0.24[0.20,0.30]
GP11	0.84(0.22)	0.73[0.67,0.79]	0.86(0.2)	0.38[0.28,0.48]	AE	0.71[0.65,0.76]		0.28[0.24,0.35]
GP12	0.67(0.32)	0.54[0.44,0.63]	0.71(0.38)	0.32[0.22,0.43]	AE	0.60[0.51,0.67]		0.40[0.33,0.49]
GP13	0.5(0.22)	0.64[0.57,0.72]	0.62(0.23)	0.46[0.37,0.55]	AE	0.70[0.63,0.75]		0.30[0.25,0.37]
GP14	14.33(4.96)	0.78[0.72,0.83]	13.6(4.22)	0.50[0.42,0.58]	ACE	0.36[0.19,0.55]	0.37[0.19,0.52]	0.27[0.22,0.33]
GP15	1.86(0.69)	0.75[0.69,0.81]	1.79(0.56)	0.55[0.47,0.63]	CE		0.66[0.61,0.70]	0.34[0.30,0.39]
GP16	3.3(0.74)	0.78[0.73,0.83]	3.45(0.65)	0.48[0.39,0.56]	ACE	0.45[0.28,0.65]	0.29[0.10,0.45]	0.26[0.21,0.32]
GP17	0.94(0.26)	0.60[0.51,0.68]	0.98(0.28)	0.32[0.22,0.42]	AE	0.62[0.54,0.68]		0.38[0.32,0.46]
GP18	9.54(3.28)	0.77[0.71,0.82]	9.68(2.86)	0.39[0.30,0.49]	AE	0.73[0.68,0.78]		0.27[0.22,0.32]
GP19	1.86(0.42)	0.54[0.45,0.64]	1.89(0.44)	0.44[0.35,0.53]	ACE	0.27[0.03,0.50]	0.29[0.10,0.47]	0.44[0.36,0.53]
GP21	1.32(0.41)	0.48[0.36,0.60]	1.31(0.42)	0.28[0.06,0.50]	AE	0.49[0.37,0.60]		0.51[0.40,0.63]
GP22	0.15(0.08)	0.27[0.14,0.39]	0.17(0.08)	0.28[0.18,0.38]	CE		0.28[0.20,0.36]	0.72[0.64,0.80]
GP23	1.65(0.53)	0.61[0.53,0.69]	1.65(0.54)	0.44[0.35,0.53]	ACE	0.37[0.14,0.59]	0.25[0.05,0.43]	0.38[0.32,0.47]
GP24	1.93(0.57)	0.53[0.43,0.62]	1.93(0.6)	0.51[0.42,0.59]	CE		0.51[0.45,0.57]	0.49[0.43,0.55]
FGS/(FG+FGS)	24.89(4.6)	0.76[0.70,0.81]	25.51(3.77)	0.41[0.31,0.50]	ACE	0.49[0.28,0.71]	0.21[0.01,0.39]	0.30[0.25,0.36]
FBGS/(FBG+FBGS)	30.66(6.62)	0.64[0.56,0.72]	30.91(6.89)	0.51[0.42,0.59]	ACE	0.32[0.12,0.52]	0.34[0.16,0.49]	0.34[0.28,0.42]
FGS/(F+FG+FGS)	18.68(4.63)	0.72[0.66,0.77]	19.2(4.07)	0.36[0.27,0.46]	AE	0.69[0.63,0.74]		0.31[0.26,0.37]
FBGS/(FB+FBG+FBGS)	21.64(5.2)	0.63[0.55,0.71]	21.86(5.57)	0.49[0.41,0.58]	ACE	0.35[0.15,0.56]	0.30[0.12,0.46]	0.35[0.28,0.42]
FG1S1/(FG1+FG1S1)	10.24(2.2)	0.76[0.70,0.81]	10.57(1.86)	0.45[0.36,0.54]	ACE	0.42[0.24,0.63]	0.29[0.09,0.45]	0.29[0.24,0.35]
FG2S1/(FG2+FG2S1+FG2S2)	37.65(6.68)	0.84[0.80,0.88]	38.9(4.76)	0.67[0.61,0.73]	CE		0.77[0.73,0.80]	0.23[0.20,0.27]
FG2S2/(FG2+FG2S1+FG2S2)	6.69(1.97)	0.64[0.57,0.72]	6.82(2.15)	0.44[0.35,0.53]	AE	0.69[0.63,0.75]		0.31[0.25,0.37]
FBG2S1/(FBG2+FBG2S1+FBG2S2)	33.2(5.86)	0.68[0.61,0.75]	33.84(4.86)	0.52[0.44,0.60]	CE		0.61[0.55,0.66]	0.39[0.34,0.45]
FBG2S2/(FBG2+FBG2S1+FBG2S2)	33.97(6.12)	0.49[0.39,0.59]	34.12(5.78)	0.48[0.39,0.57]	CE		0.48[0.41,0.54]	0.52[0.46,0.59]
FtotalS1/FtotalS2	4.34(1.36)	0.70[0.63,0.77]	4.44(1.26)	0.55[0.47,0.63]	ACE	0.19[0.01,0.37]	0.48[0.32[0.62]	0.33[0.27,0.40]
FS1/FS2	8.29(2.59)	0.74[0.68,0.80]	8.52(2.54)	0.53[0.45,0.61]	ACE	0.39[0.22,0.57]	0.34[0.18,0.49]	0.27[0.22,0.33]
FBS1/FBS2	1.01(0.25)	0.44[0.33,0.55]	1.02(0.21)	0.44[0.35,0.53]	CE		0.44[0.36,0.51]	0.56[0.49,0.64]
FBStotal/FStotal	0.28(0.1)	0.66[0.59,0.74]	0.27(0.09)	0.43[0.34,0.52]	ACE	0.23[0.02,0.46]	0.37[0.17,0.55]	0.40[0.17,0.55]
FBS1/FS1	0.16(0.06)	0.67[0.59,0.74]	0.15(0.05)	0.43[0.34,0.52]	ACE	0.39[0.18,0.61]	0.26[0.06,0.43]	0.35[0.29.0.43]
FBS1/(FS1+FBS1)	0.13(0.04)	0.69[0.62,0.76]	0.13(0.04)	0.43[0.34,0.53]	ACE	0.42[0.22,0.64]	0.25[0.05,0.41]	0.33[0.27,0.40]
FBS2/FS2	1.22(0.32)	0.67[0.60,0.74]	1.21(0.28)	0.20[0.09,0.31]	AE	0.61[0.54,0.68]		0.39[0.32,0.46]
FBS2/(FS2+FBS2)	0.54(0.06)	0.66[0.58,0.73]	0.54(0.06)	0.23[0.12,0.34]	AE	0.61[0.54,0.68]		0.39[0.32,0.46]
GP1n	0.12(0.07)	0.52[0.41,0.62]	0.13(0.08)	0.37[0.27,0.47]	AE	0.57[0.48,0.65]		0.43[0.35,0.52]
GP2n	0.6(0.34)	0.68[0.61,0.75]	0.67(0.41)	0.24[0.13,0.34]	AE	0.71[0.64,0.77]		0.29[0.23,0.36]
GP4n	24.29(7.41)	0.74[0.68,0.80]	24.21(6.4)	0.42[0.32,0.51]	ACE	0.49[0.29,0.70]	0.21[0.01,0.38]	0.30[0.25,0.36]
GP5n	0.41(0.18)	0.59[0.51,0.68]	0.51(0.21)	0.40[0.31,0.49]	AE	0.69[0.62,0.75]		0.31[0.25,0.38]
GP6n	6.65(2)	0.76[0.71,0.82]	6.67(1.9)	0.34[0.24,0.44]	AE	0.75[0.70,0.80]		0.25[0.20,0.30]
GP7n	0.68(0.3)	0.65[0.58,0.73]	0.78(0.39)	0.32[0.22,0.42]	AE	0.73[0.66,0.78]		0.27[0.22,0.34]
GP8n	23.96(2.42)	0.80[0.75,0.85]	24.33(2.45)	0.36[0.27,0.46]	AE	0.80[0.76,0.84]		0.20[0.16,0.24]

Table 2 (continued).

	MZ		DZ					
Glycan Trait	**Mean(SD)***	**ICC[95%CI]**	**Mean(SD)***	**ICC[95%CI]**	**Best model**	**A[95%CI]**	**C[95%CI]**	**E[95%CI]**
GP9n	12.59(1.72)	0.76[0.70,0.81]	12.66(1.77)	0.42[0.33,0.51]	AE	0.76[0.71,0.81]		0.24[0.19,0.29]
GP10n	7.55(1.4)	0.75[0.69,0.81]	7.58(1.44)	0.37[0.27,0.46]	AE	0.76[0.70,0.80]		0.24[0.20,0.30]
GP11n	1.06(0.24)	0.70[0.63,0.77]	1.09(0.23)	0.36[0.26,0.46]	AE	0.70[0.63,0.75]		0.30[0.25,0.37]
GP12n	0.85(0.42)	0.54[0.45,0.64]	0.92(0.53)	0.30[0.20,0.41]	AE	0.61[0.52,0.68]		0.39[0.32,0.48]
GP13n	0.63(0.29)	0.63[0.55,0.71]	0.79(0.31)	0.44[0.35,0.53]	AE	0.70[0.63,0.75]		0.30[0.25,0.37]
GP14n	18.23(6.63)	0.77[0.72,0.82]	17.39(5.83)	0.46[0.37,0.55]	ACE	0.47[0.29,0.67]	0.26[0.07,0.42]	0.27[0.22,0.33]
GP15n	2.36(0.86)	0.73[0.67,0.79]	2.28(0.74)	0.51[0.42,0.59]	ACE	0.24[0.06,0.43]	0.44[0.27,0.59]	0.32[0.26,0.38]
G0n	31.68(9.06)	0.75[0.69,0.81]	31.68(7.96)	0.40[0.30,0.49]	AE	0.72[0.66,0.76]		0.28[0.24,0.34]
G1n	45.84(2.78)	0.68[0.61,0.75]	46.44(2.79)	0.33[0.23,0.43]	AE	0.69[0.62,0.74]		0.31[0.26,0.38]
G2n	22.08(7.72)	0.76[0.70,0.81]	21.37(6.83)	0.48[0.39,0.56]	ACE	0.41[0.23,0.61]	0.31[0.12,0.46]	0.28[0.23,0.34]
Fn total	96.81(1.06)	0.55[0.46,0.64]	96.33(1.36)	0.32[0.21,0.42]	AE	0.64[0.55,0.71]		0.36[0.29,0.45]
FG0n total/G0n	98.07(0.97)	0.57[0.48,0.66]	97.89(1.08)	0.27[0.16,0.37]	AE	0.60[0.51,0.67]		0.40[0.33,0.49]
FG1n total/G1n	98.52(0.65)	0.64[0.56,0.72]	98.31(0.86)	0.33[0.23,0.43]	AE	0.72[0.65,0.78]		0.28[0.22,0.35]
FG2n total /G2n	93.04(2.25)	0.52[0.43,0.62]	91.87(2.69)	0.33[0.23,0.43]	AE	0.63[0.54,0.70]		0.37[0.30,0.46]
Fn	79.19(3.33)	0.70[0.63,0.77]	78.71(3.53)	0.36[0.26,0.46]	AE	0.72[0.66,0.77]		0.28[0.23,0.34]
FG0n/G0n	76.69(4.51)	0.71[0.65,0.78]	76.62(4.25)	0.41[0.32,0.50]	AE	0.70[0.64,0.75]		0.30[0.25,0.36]
FG1n/G1n	79.72(3.44)	0.73[0.66,0.79]	79.62(3.59)	0.37[0.28,0.47]	AE	0.74[0.68,0.79]		0.26[0.21,0.32]
FG2n/G2n	82.16(3.32)	0.56[0.47,0.65]	81.03(3.59)	0.28[0.18,0.39]	AE	0.61[0.52,0.68]		0.39[0.32,0.48]
FBn	17.62(2.98)	0.76[0.71,0.82]	17.62(2.93)	0.37[0.28,0.47]	AE	0.76[0.71,0.81]		0.24[0.19,0.29]
FBG0n/G0n	21.38(4.24)	0.75[0.69,0.81]	21.26(3.78)	0.43[0.34,0.52]	ACE	0.52[0.33,0.73]	0.20[0.01,0.37]	0.27[0.29,0.34]
FBG1n/G1n	18.8(3.29)	0.76[0.70,0.81]	18.69(3.3)	0.39[0.29,0.48]	AE	0.76[0.71,0.80]		0.24[0.20,0.29]
FBG2n/G2n	10.88(2.07)	0.70[0.63,0.77]	10.81(1.86)	0.34[0.25,0.44]	AE	0.68[0.61,0.73]		0.32[0.27,0.39]
FBn/Fn	0.22(0.05)	0.75[0.69,0.81]	0.23(0.05)	0.36[0.26,0.46]	AE	0.75[0.69,0.79]		0.25[0.21,0.31]
FBn/Fn total	18.21(3.12)	0.75[0.69,0.81]	18.3(3.12)	0.36[0.27,0.47]	AE	0.76[0.70,0.80]		0.24[0.20,0.30]s
Fn/(Bn + FBn)	4.49(0.94)	0.74[0.69,0.80]	4.42(0.93)	0.43[0.33,0.52]	AE	0.75[0.69,0.79]		0.25[0.21,0.31]
Bn/(Fn + FBn) ‰	6.54(3.06)	0.62[0.54,0.70]	8.19(3.34)	0.44[0.35,0.53]	AE	0.69[0.62,0.74]		0.31[0.26,0.38]
FBG2n/FG2n	0.13(0.03)	0.68[0.60,0.75]	0.13(0.03)	0.31[0.21,0.41]	AE	0.66[0.59,0.72]		0.34[0.28,0.41]
FBG2n /(FG2n + FBG2n)	11.71(2.3)	0.68[0.61,0.75]	11.79(2.13)	0.32[0.22,0.42]	AE	0.66[0.59,0.72]		0.34[0.28,0.41]
FG2n/(BG2n + FBG2n)	6.17(1.36)	0.70[0.64,0.77]	5.72(1.15)	0.38[0.29,0.48]	AE	0.69[0.62,0.74]		0.31[0.26,0.38]
BG2n/(FG2n + FBG2n) ‰	32.44(15.27)	0.62[0.54,0.70]	42.02(16.45)	0.43[0.34,0.52]	AE	0.73[0.67,0.78]		0.27[0.22,0.33]

Values in the three rightmost columns indicate the amount of variance attributed to the compartment of additive genetic factors (A or heritability), common/shared environmental factors (C) and unique environmental factors (E). ICC = intra-class correlation coefficient. 95% confidence intervals for both ICC and ACE are reported.

* means and SD reported are unadjusted

doi: 10.1371/journal.pone.0082558.t002

glutaminyl cyclase, which is responsible for the presence of pyroglutamyl residues in many neuroendocrine peptides.

Overall, the proportion of array-wide epigenetic associations was significantly larger (P<0.005) among glycan traits with low heritability (5/12= 41.6%) than in those showing high heritability (4/64 = 6.2%).

We also tested if the glycan traits measured could have some clinical relevance and we found that 4 of the 76 glycan traits are significantly (after adjustment for multiple tests) associated with circulating levels of triglycerides, a well known risk factor of cardiovascular risk [17] and 3 are associated with circulating levels of C-reactive protein (CRP), a well known marker of systemic inflammation [18] (Table **S2**). Although the investigation of the role of the various glycans in health and disease is beyond the scope of our study, these data suggest that the molecular mechanisms underlying IgG glycans can yield clinically relevant insights. We note that all the glycans associated with these two traits have high heritabilities and no epigenetic significant associations were found for these. However, we did not carry out a systematic study of the relationship between IgG glycans and cardiovascular or inflammatory traits which may show associations also among glycans with low heritabilities.

Table 3. Association between IgG levels and methylation probes ($P<2x10^{-6}$).

Probe	Chr	Map position(hg 18 B36)	nearest gene	Glycan	Beta	SE	*P*	h^2	$h^2>0.35$
cg13782134	16	67919362	*NRN1L*	FBS2/FS2	-0.36	0.05	$1.19x10^{-9}$	0.61	yes
cg08392591	16	89556376	*ANKRD11*	FBGS/(FB+FBG+FBGS)	0.35	0.06	$3.05x10^{-8}$	0.35	no
cg13782134	16	67919362	*NRN1L*	FBS2/(FS2+FBS2)	-0.33	0.05	$5.8x10^{-8}$	0.61	yes
cg16029957	2	37425956	*QPCT*	GP13n	-0.19	0.03	$2x10^{-7}$	0.70	yes
cg16029957	2	37425956	*QPCT*	Bn/(Fn + FBn) ‰	-0.19	0.03	$3.57x10^{-7}$	0.69	yes
cg26991199	1	24307153	*SFRS10*	GP24	0.24	0.04	$8.99x10^{-7}$	-	no
cg26991199	1	24307153	*SFRS10*	FBGS/(FB+FBG+FBGS)	0.28	0.05	$1.01x10^{-6}$	0.35	no
cg16029957	2	37425956	*QPCT*	GP13	-0.17	0.03	$1.44x10^{-6}$	0.70	yes
cg26991199	1	24307153	*SFRS10*	GP19	0.27	0.05	$1.47x10^{-6}$	0.27	no
cg08392591	16	89556376	*ANKRD11*	FBGS/(FBG+FBGS)	0.34	0.07	$1.67x10^{-6}$	0.32	no

doi: 10.1371/journal.pone.0082558.t003

Discussion

In this study we have evaluated, using a classical study design, the heritable and non-heritable component of circulating immunoglobulin G glycome composition. The study was sufficiently powered to detect with 95% probability heritabilities above 0.4 even with low-shared environmental contributions and we detected 59 glycan traits with heritabilities above this cut-off.

Our data show that variation in levels of 51 of the 76 IgG glycan traits studied is at least half heritable and only a small proportion of N-glycan traits have a low genetic contribution. Though glycans are produced in a complex biosynthetic pathway [19] and were believed to be significantly affected by many environmental factors [20], we find a high contribution of the genetic component to IgG glycome composition. Compared to GWAS study of the plasma glycome in 2705 individuals [13], the recent GWAS study of the IgG glycome in 2247 [14] has identified five times more genetic loci with genome-wide significant associations. Inter-individual variation of the IgG glycome [9] is more than three-fold larger than the inter-individual variation of the plasma glycome [10]. Large heritability and the large number of involved genetic loci suggest that the genetic regulation of the IgG glycome is stronger than the regulation of the total plasma glycome. Fine details of IgG glycan structure significantly affect function of immunoglobulins [21], thus close regulation of IgG glycosylation is required for proper function of the immune system [22]. Genetic loci with variants that affect IgG glycome composition were reported for the majority of glycans with high heritability [12-14] (Table **S3**), and some of the highly heritable glycans seem to be affected by multiple loci (e.g., FG0n/G0n associates with genetic variants in the following genes: *FUT8*, *MGAT3*, *IKZF1* and *SMARCB1-DERL3*).

In addition to the genetic contribution to the glycome, an important source of complexity and variability in IgG glycosylation is the interaction with the environment, some of which may be revealed by epigenetic changes [23]. Epigenetic silencing of *HNF1A*, a known master regulator of plasma protein fucosylation, has been shown to be associated with changes in the composition of the human plasma N-glycome [24]. In this study we find that methylation levels at other genes are also implicated in glycome composition, both in those with high heritabilities and those with a lower genetic contribution. By using a well-characterized cohort with epigenetics data available, such as TwinsUK, it was possible to integrate glycome data with other existing molecular data and adjust for confounders.

The main limitation of the present study is that due to the novelty of the glycan phenotypes we lack the replication for the epigenetic findings in an independent cohort. The fact that we find epigenome-wide significant hits on a relatively small sample suggests that epigenetic factors contribute to IgG glycan levels, although due to the lack of replication we cannot exclude false positive results. Epigenetic factors also play a role in the case of some glycans with a high heritability. However, we only found 5 significant methylation hits (mapping to two probes) for 64 glycan traits with h^2 >0.35 and 5 significant hits (mapping to two probes) for 12 glycan traits with $h^2 \leq 0.35$. The probes associated with highly heritable glycan traits were different to those associated with glycan traits with lower heritabilities.

For glycan traits with lower heritabilities the most significant probe maps to the p53 activator *ANKRD11*. The tumor suppressor p53 is known to be able to modulate innate immune gene responses. For glycan traits with high heritabilities the two hits appear related to neuroendocrine regulation, in one case directly to a glycosylation enzyme *QPCT*, in the second case to a neuritin-like protein precursor which has been implicated in neuronal survival [25]. These genes have not been previously reported to have a role in IgG glycosylation. A similar observation was recently reported in a GWA study of the IgG glycome which identified 12 genes not previously known to be involved in IgG glycosylation [14], providing further evidence that IgG glycosylation is a very complex and tightly regulated process. As more epigenetic and genetic data become available for cohorts with IgG *N*-glycan characterizations it will become possible to elucidate the molecular pathways underlying many complex traits.

Materials and Methods

Ethic statement

The study was approved by St. Thomas' Hospital Research Ethics Committee, and all twins provided informed written consent.

Study subjects

Study subjects were twins enrolled in the TwinsUK registry, a national register of adult twins. Twins were recruited as volunteers by successive media campaigns without selecting for particular diseases or traits [26]. All twin pairs recruited were of the same sex.

In this study we analysed data from 440 monozygotic and 610 dizygotic female twins with glycomics and epigenomic data available.

Isolation of IgG from human plasma

The IgG was isolated using protein G monolithic plates as described previously [9]. Briefly, 90 µl of plasma was diluted 10x with PBS, applied to the protein G plate (BIA Separations, Ljubljana, Slovenia) and instantly washed. IgGs were eluted with 1 ml of 0.1 M formic acid and neutralized with 1 M ammonium bicarbonate.

N-Glycan Release

Isolated IgG samples were dried in a vacuum centrifuge. After drying, proteins were denatured with addition of 20 µL 2% SDS (w/v) (Invitrogen, Carlsbad, CA, USA) and by incubation at 60 °C for 10 min. Subsequently, 10 µL of 4% Igepal-CA630 (Sigma-Aldrich, St. Louis, MO, USA) and 0.5 mU of PNGase F in 10 µL 5× PBS were added to the samples. The samples were incubated overnight at 37 °C for *N*-glycan release.

2-aminobenzamide labelling

The released *N*-glycans were labelled with 2-aminobenzamide (2-AB), the fluorescent dye used to make glycans visible in UPLC. The labelling mixture was freshly prepared by dissolving 2-AB (Sigma-Aldrich, St. Louis, MO, USA) in DMSO (Sigma-Aldrich, St. Louis, MO, USA) and glacial acetic acid (Merck, Darmstadt, Germany) mixture (85:15, v/v) to a final concentration of 48 mg/mL. A volume of 25 µL of labelling mixture was added to each *N*-glycan sample in the 96-well plate. Also, 25 µL of freshly prepared reducing agent solution (2-picoline borane (Sigma-Aldrich, St. Louis, MO, USA) in DMSO – concentration of 106.96 mg/ml) was added and the plate was sealed using adhesive tape. Mixing was achieved by shaking for 10 min, followed by 2 hour incubation at 65 °C. Samples (in a volume of 100 µL) were brought to 80% ACN (v/v) by adding 400 µL of ACN.

Cleaning and elution of labelled glycans using HILIC-Solid Phase Extraction (SPE)

Free label and reducing agent were removed from the samples using HILIC-SPE. An amount of 200 µL of 0.1 g/mL suspension of microcrystalline cellulose (Merck, Darmstadt, Germany) in water was applied to each well of a 0.45 µm GHP filter plate (Pall Corporation, Ann Arbor, MI, USA). Solvent was removed by application of vacuum using a vacuum manifold (Millipore Corporation, Billerica, MA, USA). All wells were prewashed using 5 × 200 µL water, followed by equilibration using 3 × 200 µL acetonitrile/water (80:20, v/v). The samples were loaded to the wells. The wells were subsequently washed 5 × using 200 µL acetonitrile/water (80:20, v/v).

Glycans were eluted 2 × with 100 µL of water and combined elutes were not dried, but either analyzed immediately by UPLC or stored at –20 °C until usage.

Hydrophilic interaction high performance liquid chromatography (HILIC)

Fluorescently labelled *N*-glycans were separated by hydrophilic interaction chromatography on a Waters Acquity UPLC instrument (Milford, MA, USA) consisting of a quaternary solvent manager, sample manager and a FLR fluorescence detector set with excitation and emission wavelengths of 330 and 420 nm, respectively. The instrument was under the control of Empower 2 software, build 2145 (Waters, Milford, MA, USA). Labeled *N*-glycans (10 µl) were separated on a Waters BEH Glycan chromatography column, 100 × 2.1 mm i.d., 1.7 µm BEH particles, with 100 mM ammonium formate, pH 4.4, as solvent A and acetonitrile as solvent B. Separation method used linear gradient of 75–62% acetonitrile (v/v) at flow rate of 0.4 ml/min in a 25 min analytical run. Samples were maintained at 5 °C before injection, and the separation temperature was 60 °C. The system was calibrated using an external standard of hydrolyzed and 2-AB labelled glucose oligomers from which the retention times for the individual glycans were converted to glucose units. Data processing was performed using an automatic processing method with a traditional integration algorithm after which each chromatogram was manually corrected to maintain the same intervals of integration for all the samples. The chromatograms obtained were all separated in the same manner into 24 peaks and the amount of glycans in each peak was expressed as % of total integrated area. In addition to 24 directly measured glycan structures, 53 derived traits were calculated as described previously [9] (see Table **S1**). These derived traits average particular glycosylation features (galactosylation, fucosylation, sialylation) across different individual glycan structures. Consequently, they are more closely related to individual enzymatic activities, and underlying genetic polymorphisms [13].

Epigenetics

DNA methylation levels were obtained using the 27k Illumina CpG methylation probe array in 127 female twins aged 32 to 80 randomly selected from the discovery cohort. QC measure were applied, as previously described [16] and 24,641 autosomal probes passed quality control and were included in the analysis. Probes were standardized to have mean zero and variance 1.

Statistical methods

Statistical analysis was carried out using Stata version 11. The R package OpenMX was used to calculate heritability.

Heritability of glycome composition was estimated using structural equation modelling to decompose the observed phenotypic variance into three latent sources of variation: additive genetic variance (A), shared/common environmental variance (C), and non-shared/unique environmental variance (E) [27] adjusting for age and batch. Additive genetic influences are indicated when monozygotic twins are more similar than dizygotic twins. The common environmental component estimates the contribution of family environment which is assumed to be equal in both MZ and DZ twin pairs [28], whereas the unique environmental component does not contribute to twin similarity, rather it estimates the effects that apply only to each individual and includes measurement error. The equal environment assumption across zygosities implies that any greater similarity between MZ twins than DZ twins is attributed to greater sharing of genetic influences.

The best fitting model (among ACE, AE, CE, and E models) was determined by removing each factor sequentially from the full model and testing the deterioration in fit of the various nested models, using the likelihood ratio test (p=0.05). In addition, the Akaike Information Criteria (AIC) was considered, with lower values indicating the most suitable model. The AIC combines the goodness of fit of a model (the discrepancy of expected to observed covariance matrixes) with its simplicity, resulting in a measure of parsimony [27]. The most parsimonious model was then used to estimate heritability, defined as the proportion of the phenotypic variation attributable to genetic factors

Random intercept logistic regression was used to test the association between whole-blood DNA methylation patterns and IgG patterns with low heritability, adjusting for age, sex, BMI, methylation chip, sample position on methylation chip, and family relatedness. Adjusting for zygosity did not change the results.

Supporting Information

Table S1. Description of the glycan codes.
(DOCX)

Table S2. List of glycans associated with circulating levels of triglycerides (log) and of C-reactive protein (Bonferroni $P<7x10^{-4}$).
(DOCX)

Table S3. List of loci associated with all glycans with high heritability ($h^2>0.55$).
(DOCX)

Author Contributions

Conceived and designed the experiments: CM MIMC TDS GL AMV. Performed the experiments: IA FV MPB OG VZ GL TK. Analyzed the data: CM MM JTB IE AMV. Contributed reagents/materials/analysis tools: IA FV MPB OG VZ GL TK. Wrote the manuscript: IA FV MPB OG VZ GL TK.

References

1. Marino K, Saldova R, Adamczyk B, Rudd PM (2012) Changes in Serum N-Glycosylation Profiles: Functional Significance and Potential for Diagnostics. In: AP RauterT Lindhorst. Carbohydrate Chemistry. Cambridge, UK: The Royal Society of Chemistry. pp. 57-93.
2. Ohtsubo K, Marth JD (2006) Glycosylation in cellular mechanisms of health and disease. Cell 126: 855-867. doi:10.1016/j.cell.2006.08.019. PubMed: 16959566.
3. Arnold JN, Wormald MR, Sim RB, Rudd PM, Dwek RA (2007) The impact of glycosylation on the biological function and structure of human immunoglobulins. Annu Rev Immunol 25: 21-50. doi:10.1146/annurev.immunol.25.022106.141702. PubMed: 17029568.
4. Taniguchi N, Honke K, Furunda M (2002) Handbook of glycosyltransferases and related genes. Tokyo: Springer Verlag.
5. Freeze HH (2006) Genetic defects in the human glycome. Nat Rev Genet 7: 537-551. doi:10.1038/nrg1894. PubMed: 16755287.
6. Nairn AV, York WS, Harris K, Hall EM, Pierce JM et al. (2008) Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. J Biol Chem 283: 17298-17313. doi:10.1074/jbc.M801964200. PubMed: 18411279.
7. Abbott KL, Nairn AV, Hall EM, Horton MB, McDonald JF et al. (2008) Focused glycomic analysis of the N-linked glycan biosynthetic pathway in ovarian cancer. Proteomics 8: 3210-3220. doi:10.1002/pmic.200800157. PubMed: 18690643.
8. Lauc G, Vojta A, Zoldoš V (2013) Epigenetic regulation of glycosylation is the quantum mechanics of biology. Biochim Biophys Acta, 1840: 65–70. PubMed: 23999089.
9. Pucić M, Knezević A, Vidic J, Adamczyk B, Novokmet M 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: 010090 21653738.
10. Knezević A, Polasek O, Gornik O, Rudan I, Campbell H et al. (2009) Variability, heritability and environmental determinants of human plasma N-glycome. J Proteome Res 8: 694-701. doi:10.1021/pr800737u. PubMed: 19035662.
11. Gornik O, Wagner J, Pucić M, Knezević A, Redzic I et al. (2009) Stability of N-glycan profiles in human plasma. Glycobiology 19: 1547-1553. doi:10.1093/glycob/cwp134. PubMed: 19726492.
12. Huffman JE, Knezevic A, Vitart V, Kattla J, Adamczyk B et al. (2011) Polymorphisms in B3GAT1, SLC9A9 and MGAT5 are associated with variation within the human plasma N-glycome of 3533 European adults. Hum Mol Genet 20: 5000-5011. doi:10.1093/hmg/ddr414. PubMed: 21908519.
13. Lauc G, Essafi A, Huffman JE, Hayward C, Knezevic A 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.
14. Lauc G, Huffman JE, Pučić M, Zgaga L, Adamczyk B et al. (2013) Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. PLoS Genet 9: e1003225. PubMed: 23382691.
15. Visscher PM (2004) Power of the classical twin design revisited. Twin Res 7: 505-512. doi:10.1375/twin.7.5.505. PubMed: 15527666.
16. Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J et al. (2012) Epigenome-Wide Scans Identify Differentially Methylated Regions for Age and Age-Related Phenotypes in a Healthy Ageing. Population - PLOS Genet 8: e1002629.
17. Ridker PM, Hennekens CH, Buring JE, Rifai N (2000) C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. N Engl J Med 342: 836-843. doi:10.1056/NEJM200003233421202. PubMed: 10733371.
18. Hokanson JE, Austin MA (1996) Plasma triglyceride level is a risk factor for cardiovascular disease independent of high-density lipoprotein cholesterol level: a meta-analysis of population-based prospective studies. J Cardiovasc Risk 3: 213-219. doi:10.1097/00043798-199604000-00014. PubMed: 8836866.
19. Lauc G, Rudan I, Campbell H, Rudd PM (2010) Complex genetic regulation of protein glycosylation. Mol Biosyst 6: 329-335. doi:10.1039/b910377e. PubMed: 20094651.

20. Lauc G, Zoldoš V (2010) Protein glycosylation--an evolutionary crossroad between genes and environment. Mol Biosyst 6: 2373-2379. doi:10.1039/c0mb00067a. PubMed: 20957246.
21. Gornik O, Pavić T, Lauc G (2012) Alternative glycosylation modulates function of IgG and other proteins - implications on evolution and disease. Biochim Biophys Acta 1820: 1318-1326. doi:10.1016/j.bbagen.2011.12.004. PubMed: 22183029.
22. Nimmerjahn F, Ravetch JV (2008) Fcgamma receptors as regulators of immune responses. Nat Rev Immunol 8: 34-47. doi:10.1038/nri2206. PubMed: 18064051.
23. Zoldoš V, Novokmet M, Bečeheli I, Lauc G (2013) Genomics and epigenomics of the human glycome. Glycoconj J 30: 41-50. doi:10.1007/s10719-012-9397-y. PubMed: 22648057.
24. Zoldoš V, Horvat T, Novokmet M, Cuenin C, Mužinić A et al. (2012) Epigenetic silencing of HNF1A associates with changes in the composition of the human plasma N-glycome. Epigenetics 7: 164-172. doi:10.4161/epi.7.2.18918. PubMed: 22395466.
25. Fujino T, Wu Z, Lin WC, Phillips MA, Nedivi E (2008) cpg15 and cpg15-2 constitute a family of activity-regulated ligands expressed differentially in the nervous system to promote neurite growth and neuronal survival. J Comp Neurol 507: 1831-1845. doi:10.1002/cne.21649. PubMed: 18265009.
26. Moayyeri A, Hammond CJ, Valdes AM, Spector TD (2013) Cohort Profile: TwinsUK and Healthy Ageing Twin Study. Int J Epidemiol, 42: 76–85. PubMed: 22253318.
27. Neale M, Cardon L (1992) Methodology for Genetic Studies of Twins and Families. Dordrecht: Kluwer Academic Publishers.
28. Kyvic K (2000) Generalisability and assumptions of twin studies. In: TD SpectorH SneiderAJ MacGregor. Advances in twin and sib-pair analysis. London: Greenwich Medical Media. pp. 67-77.

3. Glycosylation Profile of IgG in Moderate Kidney Dysfunction

CLINICAL RESEARCH www.jasn.org

Glycosylation Profile of IgG in Moderate Kidney Dysfunction

Clara Barrios,*† Jonas Zierer,*‡ Ivan Gudelj,§ Jerko Štambuk,§ Ivo Ugrina,§ Eva Rodríguez,† María José Soler,† Tamara Pavić,‖ Mirna Šimurina,‖ Toma Keser,‖ Maja Pučić-Baković,§ Massimo Mangino,* Julio Pascual,† Tim D Spector,* Gordan Lauc,§‖ and Cristina Menni*

*Department of Twin Research and Genetic Epidemiology, Kings College London, London, United Kingdom; †Department of Nephrology, Hospital del Mar, Institut Mar d'Investigacions Mediques, Barcelona, Spain; ‡Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum Munich, Germany; §Genos Glycoscience Research Laboratory, Zagreb, Croatia; and ‖University of Zagreb, Faculty of Pharmacy and Biochemistry, Zagreb, Croatia

ABSTRACT

Glycans constitute the most abundant and diverse form of the post-translational modifications, and animal studies have suggested the involvement of IgG glycosylation in mechanisms of renal damage. Here, we explored the associations between IgG glycans and renal function in 3274 individuals from the TwinsUK registry. We analyzed the correlation between renal function measured as eGFR and 76 *N*-glycan traits using linear regressions adjusted for covariates and multiple testing in the larger population. We replicated our results in 31 monozygotic twin pairs discordant for renal function. Results from both analyses were then meta-analyzed. Fourteen glycan traits were associated with renal function in the discovery sample ($P<6.5\times10^{-4}$) and remained significant after validation. Those glycan traits belong to three main glycosylation features: galactosylation, sialylation, and level of bisecting *N*-acetylglucosamine of the IgG glycans. These results show the role of IgG glycosylation in kidney function and provide novel insight into the pathophysiology of CKD and potential diagnostic and therapeutic targets.

J Am Soc Nephrol 27: ●●●–●●●, 2015. doi: 10.1681/ASN.2015010109

Chronic kidney disease affects 13% of the adult population in developed countries and it is associated with increased cardiovascular morbidity and mortality.[1,2] Though many genetic[3–5] and environmental factors (such as diabetes, hypertension and ageing)[6] are implicated in the development of kidney damage, its physiopathology is still not fully understood. Heritability estimates for CKD range between 0.33 and 0.41[7,8] and despite the discovery of several important genetic associations, these loci collectively account for only 1.4% of the variation in eGFR.[5] This suggests that epigenetic or post-transcriptional factors may be playing an important role in renal damage.

Glycosylation is the most abundant and diverse form of post-transcriptional modification and participates in every physiologic process.[9]

Immunoglobulin G is an excellent glycoprotein model as its glycosylation is well defined and many important functional effects of alternative IgG glycosylation have been described.[10] *N*-glycans attached to the conserved asparagine 297 in the Fc part of IgG are important modulators of IgG effector functions.[11] For example, glycosylation acts as a switch between pro- and anti-inflammatory IgG functionality. Malfunction of this system is associated with different inflammatory and autoimmune

Received January 30, 2015. Accepted June 5, 2015.

C.B., J.Z., G.L., C.M. and I.G. contributed equally to this work.

Published online ahead of print. Publication date available at www.jasn.org.

Correspondence: Dr. Clara Barrios, Department of Nephrology, Hospital del Mar, Institut Mar d'Investigacions Mediques, Barcelona, Spain, or Dr. Cristina Menni, Department of Twins Research and Genetic Epidemiology, Kings College London, London, UK. E-mail: cristina.menni@kcl.ac.uk or CBarrios@parcdesalutmar.cat.

diseases such as SLE,[12] rheumatoid arthritis, inflammatory bowel diseases,[13,14] cancer[15,16] and AIDS.[17] Furthermore, it has been shown that inflammation pathways play a key role in endothelial and kidney damage.[18,19] Indeed, the activation of inflammatory pathways and subsequent fibrosis are hallmark of renal injury.[20,21] Different IgG glycosylation profiles may provide an at-risk phenotype to the development of renal damage.

Animal models highlighted the potential role of IgG glycosylation in the pathophysiologic mechanism involved in renal damage. Indeed studies have shown that modulation of ANCA IgG glycosylation reduces its pathogenicity in mouse ANCA-associated GN.[21] Also, IgG Fcγ receptor deficiency was found to be renoprotective in a mouse model of diabetic nephropathy.[20] Human studies suggest that aberrant glycosylation of the IgA1 is implicated in the deposit and formation of the immunocomplex IgA–IgG in patients with IgA nephropathy.[22,23]

However, no human studies investigated the role of the IgG glycosylation profiles in the onset of CKD.

The aim of this study is to investigate the potential role of IgG glycosylation in kidney function, by analyzing IgG glycome composition in a large population-based cohort from the UK. As glycans are associated with many factors including genes,[24] we validate our significant results in an independent population of identical twins discordant for renal diseases.

RESULTS

Levels of 76 IgG glycans (24 directly measured and 52 derived traits) (Supplemental Figure 1) were obtained in 3274 individuals with different eGFR from the TwinsUK population (age range: 18–87 years). The demographic characteristics of the study populations are presented in Table 1. We identified 31 monozygotic (MZ) twin pairs discordant for the renal phenotype (difference in eGFR>15 mL/min per 1.73 m^2).

We first ran the linear regressions in the discovery population adjusting for age, sex, body mass index (BMI), diabetes, hypertension, glycan analysis batch and family relatedness, excluding the MZ discordant twins. We controlled for multiple testing using Bonferroni correction ($P<6.5\times10^{-4}$; 0.05/76 glycan traits). This identified 14 glycans significantly associated with eGFR; six glycans were positively associated with eGFR, while eight were negatively associated (Table 2, Supplemental Table 1). To ensure that sexual hormones did not affect our results, we ran the same linear regression analysis including menopause as a covariate and our results were unchanged.

We then assessed whether these associations with renal function were robust by testing an independent group of MZ twins discordant for renal disease. The regression coefficients were in the same direction in both analyses (discordant identical twins and the rest of the population). We then combined the results using inverse-variance fixed effect meta-analysis. All 14 glycans remained Bonferroni significant (Table 2). As depicted in Figure 1 and Table 2, the 14 significant glycan traits fell into three particular glycosylation features: galactosylation, sialylation and the level of bisecting *N*-acetylglucosamine (GlcNAc) of the IgG glycans.

We observed a decrease in agalactosylated glycans: A2 (GP2 and $GP2^n$) and FA2B (GP6 and $GP6^n$) glycan structures and derived trait $G0^n$, which combines all agalactosylated structures. Conversely, glycan with galactose on both antennae, FA2G2 (GP14 and $GP14^n$), and the $G2^n$ derived trait, representing the percentage of digalactosylated structures in neutral IgG glycans, increased in parallel with the eGFR. The same pattern was observed in the MZ discordant pairs. As for sialylation, the major sialylated glycan, FA2G2S1 (GP18) and the percentage of sialylated structures without bisecting GlcNAc (represented by the ratio FGS/[F+FG+FGS]) increased with eGFR.

The level of bisecting GlcNAc in sialylated IgG glycans represented by three ratios, FBS^{total}/FS^{total}, FBS1/FS1, and FBS1/(FS1+FBS1), as well as in digalactosylated neutral gG glycans ($FG2^n/[BG2^n+FBG2^n]$) were found to be inversely associated with eGFR.

To reinforce our findings we searched for associations in an independent population with more severe renal phenotype (eGFR<30 mL/min per 1.73 m^2). Eight twins, mean aged 65.0 (range 42.2–75.5 years) with CKD stage 4/5 (mean eGFR 24.7 [range 8.0–27.3]) were compared with their age-matched cotwin with eGFR>30 ml/min per 1.73 m^2. As depicted in Figure 2, IgG glycans profiles follow the same patterns as were observed in the discovery population with the worsening of the renal function.

To determine whether the findings were restricted to IgG or to a more general change in glycosylation of multiple proteins, we searched for association between total plasma glycome[25,26] and eGFR in a subset 426 individuals (eGFR, mL/min/1.73 m^2: 78.95±16.00). We found no difference in plasma glycosylation, suggesting that the effects we see here are likely direct effects of IgG glycosylation. However, the lack of association might also be due to power issues and so further study on larger sample size is needed to test this (Supplementary Table 2).

Table 1. General characteristics of the study population

	Discovery Population	MZ Discordant Twins
Sample size, *n*	3212	62
Age, years	52.67±14.15	55.45±12.2
MZ:DZ:singletons	506:1772:934	62:0:0
Female, *n* (%)	3050 (94.9)	60 (96.7)
BMI, kg/m^2	25.95±4.65	25.64±5.65
Creatinine, *mg/ml*	0.83±0.15	0.75±0.10
eGFR, mL/min per 1.73 m^2	84.15±17.02	88.52±9.91
CKD (eGFR≤60), *n* (%)	294 (9.15)	1 (1.6)
Type II diabetes, *n* (%)	72 (2.2)	4 (6.4)
Hypertension, *n* (%)	705 (21.9)	18 (29.0)

CKD eGFR estimated using Chronic Kidney Disease Epidemiology Collaboration equation. Values for categorical variables are given as *n* (%); values for continuous variable as mean (±SD). MZ:DZ, monozygotic:dizygotic.

Table 2. Glycan traits significantly associated with eGFR in the discovery, validation, and meta-analysis

Glycan	Description	h^{2a}	Discovery		MZ Discordant	Fixed effect meta-analysis	
			β (95% CI)	*P* value	β (95% CI)	β (95% CI)	*P* value
GP18	The percentage of FA2G2S1 glycan in total IgG glycans	0.73	1.48 (0.89 to 2.07)	8.60×10^{-7}	0.59 (−2.23 to 3.41)	4.23 (2.38 to 7.52)	9.51×10^{-7}
GP14	The percentage of FA2G2 glycan in total IgG glycans	0.36	1.46 (0.85 to 2.07)	2.92×10^{-6}	1.33 (−1.81 to 4.48)	4.29 (2.35 to 7.81)	2.04×10^{-6}
$GP6^{n}$	The percentage of FA2B glycan in total neutral IgG glycans (GP^{n})	0.75	−1.39 (−1.98 to −0.80)	3.56×10^{-6}	−0.84 (−3.44 to 1.76)	0.26 (0.14 to 0.45)	3.16×10^{-6}
$GP14^{n}$	The percentage of FA2G2 glycan in total neutral IgG glycans (GP^{n})	0.47	1.29 (0.68 to 1.90)	3.06×10^{-5}	1.99 (−1.70 to 5.67)	3.70 (2.03 to 6.73)	1.82×10^{-5}
FBS1/FS1	Ratio of fucosylated monosialylated structures with and without bisecting GlcNAc	0.39	−1.12 (−1.65 to −0.59)	3.48×10^{-5}	−0.58 (−3.16 to 1.99)	0.33 (0.20 to 0.56)	3.42×10^{-5}
FBS1/(FS1+FBS1)	The incidence of bisecting GlcNAc in all fucosylated monosialylated structures in total IgG glycans	0.42	−1.10 (−1.63 to −0.57)	4.63×10^{-5}	−0.60 (−3.14 to 1.95)	0.34 (0.20 to 0.57)	4.46×10^{-5}
$G2^{n}$	The percentage of digalactosylated structures in total neutral IgG glycans	0.41	1.20 (0.60 to 1.80)	8.81×10^{-5}	1.98 (−1.83 to 5.78)	3.38 (1.87 to 6.10)	5.53×10^{-5}
GP6	The percentage of FA2B glycan in total IgG glycans	0.75	−1.14 (−1.71 to −0.57)	8.90×10^{-5}	−1.01 (−3.78 to 1.76)	0.32 (0.18 to 0.56)	6.84×10^{-5}
FBS^{total}/FS^{total}	Ratio of all fucosylated sialylated structures with and without bisecting GlcNAc	0.23	−1.07 (−1.60 to −0.54)	8.21×10^{-5}	−0.30 (−2.84 to 2.23)	0.36 (0.21 to 0.60)	9.52×10^{-5}
$G0^{n}$	The percentage of agalactosylated structures in total neutral IgG glycans	0.72	−1.16 (−1.76 to −0.56)	1.52×10^{-4}	−1.20 (−4.71 to 2.31)	0.31 (0.17 to 0.57)	1.20×10^{-4}
$GP2^{n}$	The percentage of A2 glycan in total neutral IgG glycans (GP^{n})	0.71	−0.91 (−1.42 to −0.40)	5.02×10^{-4}	−2.00 (−4.66 to 0.67)	0.39 (0.23 to 0.64)	2.20×10^{-4}
GP2	The percentage of A2 glycan in total IgG glycans	0.72	−0.90 (−1.42 to −0.38)	6.28×10^{-4}	−2.33 (−5.13 to 0.47)	0.39 (0.23 to 0.64)	2.55×10^{-4}
FGS/(F+FG+FGS)	The percentage of sialylation of all fucosylated structures without bisecting GlcNAc in total IgG glycans	0.69	1.01 (0.46 to 1.56)	2.96×10^{-4}	0.57 (−2.21 to 3.35)	2.71 (1.58 to 4.64)	2.85×10^{-4}
$FG2^{n}/(BG2^{n} + FBG2^{n})$	Ratio of fucosylated digalactosylated nonbisecting GlcNAc structures and all digalactosylated structures with bisecting GlcNAc	*0.66*	0.91 (0.38 to 1.44)	7.32×10^{-4}	0.93 (−1.59 to 3.44)	2.49 (1.48 to 4.19)	5.54×10^{-4}

[a]Estimates of heritability (h^2) come from Menni *et al. Plos One 2013.*[24]

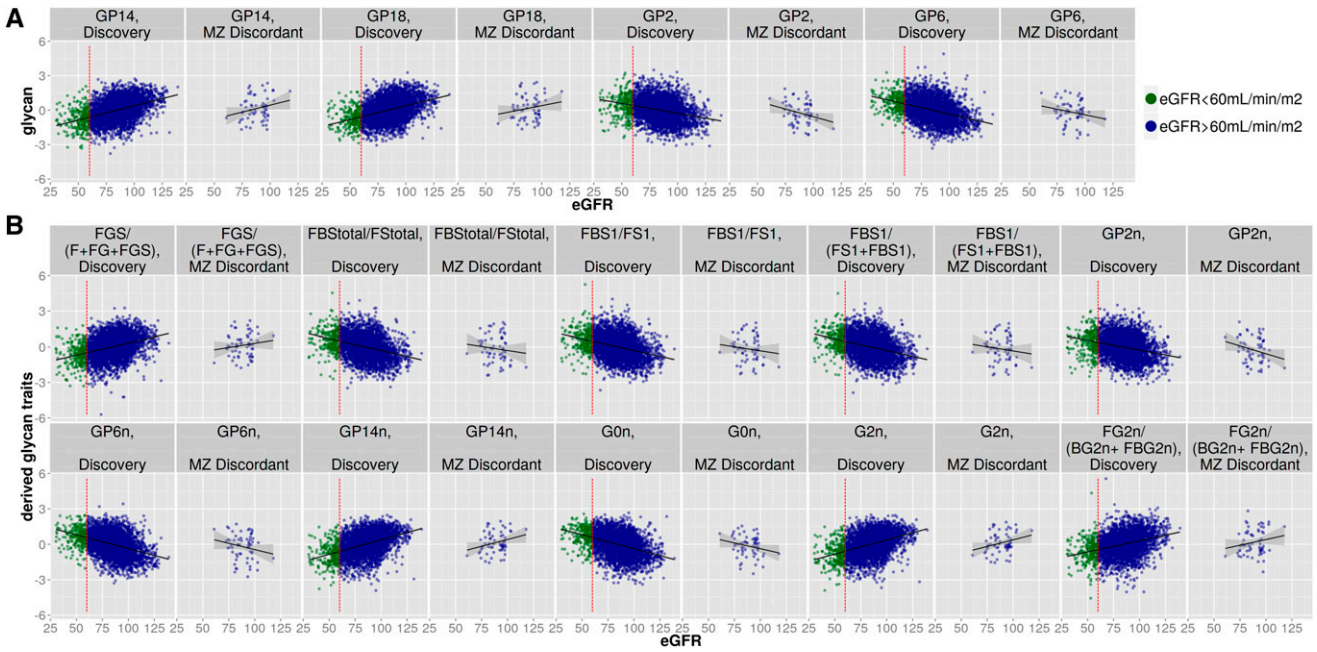


Figure 1. Correlation of IgG glycosylation and eGFR in the discovery and MZ discordant populations. (A) Directly measured glycan structures. (B) Derived traits that measure sialylation, galactosylation, and bisecting GlcNAc.

Finally, we assessed whether glycan profiles could improve the prediction of the CKD status (as per Guidelines, CKD cases have eGFR<60 mL/min per 1.73 m^2) beyond that achieved with age and sex. In the discriminative model only the four main glycans (GP2, GP6, GP14, and GP18) were included. The predictive ability for CKD status, as measured by the area under the curve was 0.87 (95% confidence interval [95% CI], 0.85 to 0.89) for clinical parameters alone, 0.81 (95% CI, 0.78 to 0.84) for glycans alone, and 0.88 (95% CI, 0.86 to 0.90) for the model incorporating a combination of glycans and clinical parameters (P=0.23) (Supplemental Figure 2).

DISCUSSION

This is the first study to investigate the potential role of IgG glycosylation in kidney function. We identified 14 IgG glycan traits with high statistical significance associated with eGFR and validated them in an independent subset of MZ twins discordant for renal disease. Moreover we see the same pattern in a small independent sample with a more extreme renal dysfunction.

The glycans identified fall into three principal glycan traits.

Galactosylation of IgG

Decreased IgG galactosylation has been found to be associated with rheumatoid arthritis[27] as well as with several autoimmune and inflammatory diseases[16] and with chronologic and biologic age.[28] The decrease in galactosylation is not disease-specific, but a general phenomenon that is associated with decreased immunosuppressive and anti-inflammatory potential of circulating IgG. We observed a higher risk of CKD in subjects with agalactosylated glycans (GP2, GP6, and $G0^n$) and lower in those with galactosylated IgG (GP14 and $G2^n$). Lack of terminal galactose activates complement cascade and makes IgG pro-inflammatory, whereas the addition of galactose decreases its inflammatory potential.[29,30] Hence, the IgG galactosylation pattern observed in our population supports the theory that complement activation/dysregulation is crucial in renal damage.[31] It is not clear whether IgG galactosylation is a consequence or an individual predisposition for a disease. The heritability of galactosylated glycans was very high,[24] indicating that galactosylation could partly be genetically predetermined. This hypothesis is further supported by the fact that in rheumatoid arthritis, the decrease in IgG galactosylation was observed up to several years before the onset of the disease.[32–35]

Sialylation

Further extension of IgG glycans by the addition of sialic acid dramatically changes the physiologic role of IgG, converting it from a proinflammatory into an anti-inflammatory agent.[36,37] This relatively small fraction of sialylated IgG is believed to be responsible for the immunosuppressive activity of intravenously administered immunoglobulins.[38] Approximately 50% of IgG glycans are not sialylated and are proinflammatory.[39] However, the terminal α2,6-sialylation of IgG glycans decreases the ability of IgG to bind Fcγ receptors (FcγRs), which increases expression of inhibitory FcγRIIB and is anti-inflammatory.[40] Contrary to changes in galactosylation, the significant changes in sialylation have not been associated

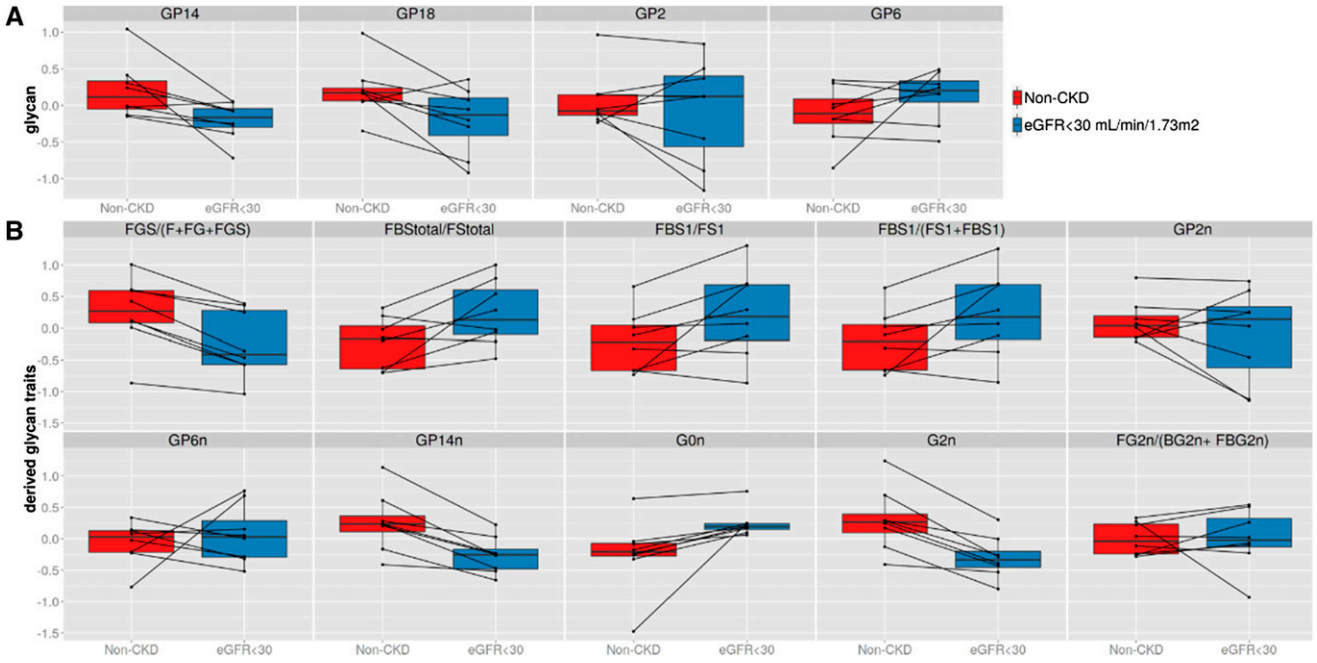


Figure 2. IgG glycan profiles in eight pairs of twins discordant for renal function. Comparisons between each pair of twins where one has extreme renal phenotype (eGFR<30 mL/min per 1.73 m^2) versus non-CKD. (A) Directly measured glycans structures. (B) Derived traits that measure sialylation, galactosylation, and bisecting GlcNAc. Results are in line with those observed in the discovery population (Figure 1).

with other diseases. Recently, some of us found that major sialylated glycans (GP16, GP18, and GP23) were significantly decreased in patients with SLE (F. Vučković *et al.*, submitted for publication). In our population, the major sialylated glycan, FA2G2S1 (GP18), and the ratio FGS/(F+FG+FGS), which represents the percentage of sialylated structures without bisecting GlcNAc in total IgG glycans, were decreased in patients with CKD (green dots in Figure 1). These sialylated glycan traits displayed a protective independent risk for CKD.

Bisecting *N*-Acetylglucosamine and Core Fucosylation of IgG

Another feature is the role of core fucose in the modulation of antibody-dependent cellular cytotoxicity.[41] On average, 95% of the IgG population is core fucosylated[42]; hence, most of the immunoglobulins have a "safety switch", which prevents them from antibody-dependent cellular cytotoxicity. IgG-containing glycans that lack core fucose have 100-fold higher affinity to the FcγRIIIa and are therefore much more efficient than fucosylated glycoforms.[43] We have observed a significant and independent decreased risk of CKD when sialylated and core fucosylated glycans did not have bisecting GlcNAc; and in contrast, lower eGFR if those glycans contained bisecting GlcNAc (FBS^{total}/FS^{total}, FBS1/FS1, and FBS1/[FS1+FBS1]). Also for neutral digalactosylated glycans, when there is less of these glycans with bisecting GlcNAc, the ratio $FG2^n/(BG2^n+FBG2^n)$ is higher and this is positively associated with eGFR. The presence of bisecting GlcNAc was always associated with a higher risk of CKD.

It is not clear how the modulation of antibody-dependent cellular cytotoxicity could affect the renal damage in the onset of a nonautoimmune CKD. Studies in experimental animals have reported that modifications in the Fcγ receptor can diminish renal damage in a well known autoimmune disease, ANCA-related GN, as well as in diabetic nephropathy.[20,21] On the other hand, renal fibrosis is the common pathway of many kidney diseases and leads to progressive renal failure; natural killer cells have been linked with this process in different organ systems.[11]

Notably, glycan traits associated with lower eGFR have on average a higher heritability (Table 2). For example, the agalactosylated IgG glycans we found associated with lower eGFR, have a high heritability, ranging from 0.72 to 0.75, whereas galactosylated glycans GP14 and $G2^n$ derived trait have a low heritability (0.36 and 0.41, respectively).[24] The highly heritable glycans associated with eGFR, have been previously associated with different genes.[12] However, there is as yet no overlap with genes previously reported in CKD genome-wide association studies.[5] Our findings may indicate a new approach to deeper understanding of the contribution of genetics in IgG glycosylation and kidney damage.

Although the identified glycans do not predict incident CKD (defined as eGFR<60 mL/min per 1.73 m^2) more accurately than clinical parameter, their inclusion in the models improves the incident CKD risk prediction. These glycans may be more sensitive to earlier stages of reduced renal function, as the eGFR-defined onset of CKD occurs only after half of the

kidneys' filtration ability has been lost. Longitudinal studies could help to address this hypothesis.

The present study has several strengths. First, we employed a two-stage design (discovery and independent replication with stringent *P* values), so minimizing the risk of false positive findings. Second, we used identical twins discordant for renal function in the validation analysis. Glycan levels may be influenced by many factors including genetics, age and environment.[12] As identical twins share 100% of their genetic makeup, and are matched perfectly for age, gender, social class, *etc.*, we were able to validate the role of IgG on renal function; isolating the nongenetic contribution. These data help us to understand the complex interplay between genetic and nongenetic influences that determine renal function.

We note some study limitations. First, there is a female predominance in our study sample (95% of the individuals are, for historical reasons, women). Second, our population being volunteers is slightly healthier than average with a lower rate of diabetes and results might not be generalizable to more severe diabetes populations. Third, the cross-sectional nature of our data does not allow us to draw conclusions as to whether the glycans identified are causative of kidney function decline or merely correlated with it. Finally, we cannot provide reliable estimates as to what proportions of the identified glycans were from Fc and from Fab, respectively. However, in a small pilot of Fc-glycopeptides by nano-liquid chromatography tandem mass spectrometry[39] on 96 representative age-matched individuals from the extremes of the eGFR distribution, we find the same direction of effect with renal function for all but one. This suggests that our initial observations mostly come from the Fc glycans (Supplemental Table 3).

Our results highlight the promising role of glycomics in renal studies. Uncovering this relationship by extending the research with clinical subsets and longitudinal data would help to identify further novel markers that would be potentially useful to detect at-risk patients, in the early stages of CKD. These results open new avenues to our understanding of renal damage and encourage further studies in populations with more severe CKD and proteinuria information, as well as studies comparing patients with autoimmune CKD with patients whose CKD is due to other etiologies. Moreover, this would help to gain additional insights into the pathophysiology of CKD and potential therapeutic targets.

CONCISE METHODS

Study Subjects

Study subjects were twins enrolled in the TwinsUK registry, a national register of adult twins. Twins were recruited as volunteers by successive media campaigns without selecting for particular diseases or traits.[44] In this study we analyzed data from 3274 individuals with glycomics and creatinine data available. The study was approved by St. Thomas' Hospital Research Ethics Committee, and all twins provided informed written consent.

Phenotype Definitions

Data relevant to the present study include BMI (body weight in kilograms divided by the square of height in square meters), type II diabetes (defined if fasting glucose $\geq$7 mmol/L or physician's letter confirming diagnosis) and hypertension. Renal parameters; eGFR was calculated from standard creatinine using the Chronic Kidney Disease Epidemiology Collaboration equation.[45] CKD was defined as an eGFR$<$60 ml/min per 1.73 m^2 according to the current Kidney Disease Outcome Quality Initiative (K/DOQI) guidelines.[46] MZ pairs were considered discordant for renal function if one twin had an eGFR$\geq$90 and the other had eGFR$\leq$90 mL/min per 1.73 m^2 and the difference between their eGFR levels was $>$15 ml/min per 1.73 m^2.

Analysis of IgG Glycans

Isolation of IgG from Human Plasma

The IgG was isolated using protein G monolithic plates (BIA Separations, Ajdovščina, Slovenia) as described previously.[42]

Glycan Release and Labeling

Glycan release and labeling were performed essentially as previously described.[24,42] Briefly, dried IgG was denatured with 2% SDS (wt/vol) and *N*-glycans were released by digestion with PNGase F (ProZyme, Hayward, CA). After deglycosylation, *N*-glycans were labeled with 2-AB fluorescent dye. Free label and reducing agent were removed from the samples using hydrophilic interaction chromatography–solid-phase extraction.

Hydrophilic Interaction Chromatography-UPLC

Fluorescently labeled *N*-glycans were separated by hydrophilic interaction chromatography on a Waters Acquity UPLC instrument (Waters, Milford, MA) as described previously.[42] Data processing was performed using an automatic processing method with a traditional integration algorithm after which each chromatogram was manually corrected to maintain the same intervals of integration for all the samples. The chromatograms were all separated in the same manner into 24 peaks and the amount of glycans in each peak was expressed as a percentage of the total integrated area. In addition to 24 directly measured glycan structures, 52 derived traits were calculated. These derived traits average particular glycosylation features (galactosylation, fucosylation, bisecting GlcNAc, and sialylation) (Supplemental Figure 1, Table 1).

Statistical Analysis

Statistical analysis was carried out using Stata version 12 and R (version 3.1.2) and visualized using the ggplot2 package.

Glycans were globally normalized and log transformed using the right-skewness of their distributions. To remove experimental biases, all measurements were adjusted for batch and run-day effects using ComBat (R-package sva). Derived glycan traits were calculated using normalized and batch-corrected glycan measurements (exponential of batch corrected measurements). All variables were centered and scaled to have mean 0 and standard deviation 1. Outliers (more than 6SD from the mean) were excluded from the analysis.

Association analyses between eGFR and glycan traits were performed using random intercept linear regressions adjusting for age, sex, BMI, diabetes, hypertension, and family relatedness as

random effect. We used a conservative Bonferroni correction to account for multiple testing assuming 76 independent tests as suggested by Pucic *et al.*,[42] so giving a significant threshold of ($P<6.5\times10^{4}$; 0.05/76). The Bonferroni-significant eGFR glycan associations were replicated in the previously excluded group of MZ discordant twins using the same model. Paired *t*-tests were used to evaluate the association with incident CKD in an independent subset of twins where one co-twin had a significant decline in renal function.

To assess how glycans can improve the prediction of CKD (eGFR$<$60 ml/min per 1.73 m^2), three Least Absolute Shrinkage and Selection Operator regression models were created (R package glmnet): The first one using only clinical parameters; age, sex, type II diabetes, and hypertension, to predict CKD, the second using the set of original glycan traits, which were found to be Bonferroni significant before (GP2, GP6, GP14, GP18), and the last one using both glycans and clinical parameters. The quality of all three models was assessed using a ten-fold cross-validation. The regularization parameter λ was trained separately for each fold using a nested cross-validation. Receiver operating characteristic curves (and particularly the area under the curves) were calculated for each fold and averages and confidence intervals were reported.

ACKNOWLEDGMENTS

TwinsUK was funded by the Wellcome Trust, European Community's Seventh Framework Programme (FP7/2007-2013).

The study also receives support from the National Institute for Health Research (NIHR) Clinical Research Facility at Guy's & St. Thomas' NHS Foundation Trust and NIHR Biomedical Research Centre based at Guy's and St. Thomas' NHS Foundation Trust and King's College London.

Tim Spector is holder of an ERC Advanced Principal Investigator award.

Clara Barrios is supported by a grant from the Spanish Society of Nephrology and RedinRen RD12/0021/0024.

Julio Pascual, María José Soler, and Eva Rodríguez are supported by grants from Institute Carlos III and RedinRen.

Glycan analysis was partly funded by European Commission FP7 projects IBD-BIOM (contract #305479), HighGlycan (contract #278535), MIMOmics (contract #305280), HTP-GlycoMet (contract #324400), and IntegraLife (contract #315997) grants.

DISCLOSURES

Gordan Lauc is founder and owner of Genos, a private research organization that specializes in high-throughput glycomic analysis and has several patents in this field.

REFERENCES

1. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu CY: Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med* 351: 1296–1305, 2004
2. Coresh J, Selvin E, Stevens LA, Manzi J, Kusek JW, Eggers P, Van Lente F, Levey AS: Prevalence of chronic kidney disease in the United States. *JAMA* 298: 2038–2047, 2007
3. Pattaro C, Köttgen A, Teumer A, Garnaas M, Böger CA, Fuchsberger C, Olden M, Chen MH, Tin A, Taliun D, Li M, Gao X, Gorski M, Yang Q, Hundertmark C, Foster MC, O'Seaghdha CM, Glazer N, Isaacs A, Liu CT, Smith AV, O'Connell JR, Struchalin M, Tanaka T, Li G, Johnson AD, Gierman HJ, Feitosa M, Hwang SJ, Atkinson EJ, Lohman K, Cornelis MC, Johansson Å, Tönjes A, Dehghan A, Chouraki V, Holliday EG, Sorice R, Kutalik Z, Lehtimäki T, Esko T, Deshmukh H, Ulivi S, Chu AY, Murgia F, Trompet S, Imboden M, Kollerits B, Pistis G, Harris TB, Launer LJ, Aspelund T, Eiriksdottir G, Mitchell BD, Boerwinkle E, Schmidt H, Cavalieri M, Rao M, Hu FB, Demirkan A, Oostra BA, de Andrade M, Turner ST, Ding J, Andrews JS, Freedman BI, Koenig W, Illig T, Döring A, Wichmann HE, Kolcic I, Zemunik T, Boban M, Minelli C, Wheeler HE, Igl W, Zaboli G, Wild SH, Wright AF, Campbell H, Ellinghaus D, Nöthlings U, Jacobs G, Biffar R, Endlich K, Ernst F, Homuth G, Kroemer HK, Nauck M, Stracke S, Völker U, Völzke H, Kovacs P, Stumvoll M, Mägi R, Hofman A, Uitterlinden AG, Rivadeneira F, Aulchenko YS, Polasek O, Hastie N, Vitart V, Helmer C, Wang JJ, Ruggiero D, Bergmann S, Kähönen M, Viikari J, Nikopensius T, Province M, Ketkar S, Colhoun H, Doney A, Robino A, Giulianini F, Krämer BK, Portas L, Ford I, Buckley BM, Adam M, Thun GA, Paulweber B, Haun M, Sala C, Metzger M, Mitchell P, Ciullo M, Kim SK, Vollenweider P, Raitakari O, Metspalu A, Palmer C, Gasparini P, Pirastu M, Jukema JW, Probst-Hensch NM, Kronenberg F, Toniolo D, Gudnason V, Shuldiner AR, Coresh J, Schmidt R, Ferrucci L, Siscovick DS, van Duijn CM, Borecki I, Kardia SL, Liu Y, Curhan GC, Rudan I, Gyllensten U, Wilson JF, Franke A, Pramstaller PP, Rettig R, Prokopenko I, Witteman JC, Hayward C, Ridker P, Parsa A, Bochud M, Heid IM, Goessling W, Chasman DI, Kao WH, Fox CS; CARDIoGRAM Consortium; ICBP Consortium; CARe Consortium; Wellcome Trust Case Control Consortium 2 (WTCCC2): Genome-wide association and functional follow-up reveals new loci for kidney function. *PLoS Genet* 8: e1002584, 2012
4. Okada Y, Sim X, Go MJ, Wu JY, Gu D, Takeuchi F, Takahashi A, Maeda S, Tsunoda T, Chen P, Lim SC, Wong TY, Liu J, Young TL, Aung T, Seielstad M, Teo YY, Kim YJ, Lee JY, Han BG, Kang D, Chen CH, Tsai FJ, Chang LC, Fann SJ, Mei H, Rao DC, Hixson JE, Chen S, Katsuya T, Isono M, Ogihara T, Chambers JC, Zhang W, Kooner JS, Albrecht E, Yamamoto K, Kubo M, Nakamura Y, Kamatani N, Kato N, He J, Chen YT, Cho YS, Tai ES, Tanaka T; KidneyGen Consortium; CKDGen Consortium; GUGC consortium: Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. *Nat Genet* 44: 904–909, 2012
5. Köttgen A, Pattaro C, Böger CA, Fuchsberger C, Olden M, Glazer NL, Parsa A, Gao X, Yang Q, Smith AV, O'Connell JR, Li M, Schmidt H, Tanaka T, Isaacs A, Ketkar S, Hwang SJ, Johnson AD, Dehghan A, Teumer A, Paré G, Atkinson EJ, Zeller T, Lohman K, Cornelis MC, Probst-Hensch NM, Kronenberg F, Tönjes A, Hayward C, Aspelund T, Eiriksdottir G, Launer LJ, Harris TB, Rampersaud E, Mitchell BD, Arking DE, Boerwinkle E, Struchalin M, Cavalieri M, Singleton A, Giallauria F, Metter J, de Boer IH, Haritunians T, Lumley T, Siscovick D, Psaty BM, Zillikens MC, Oostra BA, Feitosa M, Province M, de Andrade M, Turner ST, Schillert A, Ziegler A, Wild PS, Schnabel RB, Wilde S, Munzel TF, Leak TS, Illig T, Klopp N, Meisinger C, Wichmann HE, Koenig W, Zgaga L, Zemunik T, Kolcic I, Minelli C, Hu FB, Johansson A, Igl W, Zaboli G, Wild SH, Wright AF, Campbell H, Ellinghaus D, Schreiber S, Aulchenko YS, Felix JF, Rivadeneira F, Uitterlinden AG, Hofman A, Imboden M, Nitsch D, Brandstätter A, Kollerits B, Kedenko L, Mägi R, Stumvoll M, Kovacs P, Boban M, Campbell S, Endlich K, Völzke H, Kroemer HK, Nauck M, Völker U, Polasek O, Vitart V, Badola S, Parker AN, Ridker PM, Kardia SL, Blankenberg S, Liu Y, Curhan GC, Franke A, Rochat T, Paulweber B, Prokopenko I, Wang W, Gudnason V, Shuldiner AR, Coresh J, Schmidt R, Ferrucci L, Shlipak MG, van Duijn CM, Borecki I, Krämer BK, Rudan I, Gyllensten U, Wilson JF, Witteman JC, Pramstaller

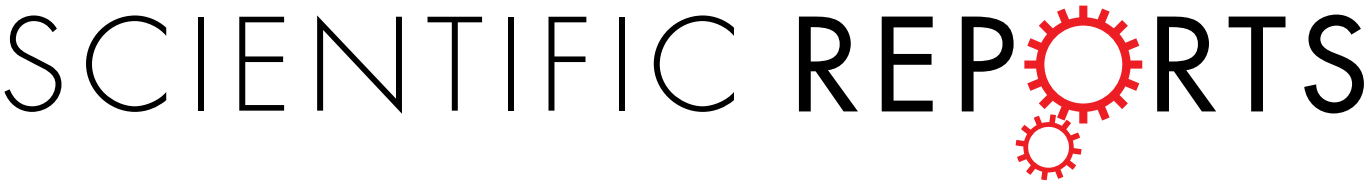
PP, Rettig R, Hastie N, Chasman DI, Kao WH, Heid IM, Fox CS: New loci associated with kidney function and chronic kidney disease. *Nat Genet* 42: 376–384, 2010
6. Meguid El Nahas A, Bello AK: Chronic kidney disease: the global challenge. *Lancet* 365: 331–340, 2005
7. Bochud M, Elston RC, Maillard M, Bovet P, Schild L, Shamlaye C, Burnier M: Heritability of renal function in hypertensive families of African descent in the Seychelles (Indian Ocean). *Kidney Int* 67: 61–69, 2005
8. Langefeld CD, Beck SR, Bowden DW, Rich SS, Wagenknecht LE, Freedman BI: Heritability of GFR and albuminuria in Caucasians with type 2 diabetes mellitus. *Am J Kidney Dis* 43: 796–800, 2004
9. National Research Council: Committee on Assessing the Importance and Impact of Glycomics and Glycosciences. Transforming Glycoscience: A Roadmap for the Future, Washington, D.C., The National Academies Press, 2012
10. Gornik O, Pavić T, Lauc G: Alternative glycosylation modulates function of IgG and other proteins - implications on evolution and disease. *Biochim Biophys Acta* 1820: 1318–1326, 2012
11. Jang HR, Rabb H: Immune cells in experimental acute kidney injury. *Nat Rev Nephrol* 11: 88–101, 2015
12. Lauc G, Huffman JE, Pučić M, Zgaga L, Adamczyk B, Mužinić A, Novokmet M, Polašek O, Gornik O, Krištić J, Keser T, Vitart V, Scheijen B, Uh HW, Molokhia M, Patrick AL, McKeigue P, Kolčić I, Lukić IK, Swann O, van Leeuwen FN, Ruhaak LR, Houwing-Duistermaat JJ, Slagboom PE, Beekman M, de Craen AJ, Deelder AM, Zeng Q, Wang W, Hastie ND, Gyllensten U, Wilson JF, Wuhrer M, Wright AF, Rudd PM, Hayward C, Aulchenko Y, Campbell H, Rudan I: Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. *PLoS Genet* 9: e1003225, 2013
13. Go MF, Schrohenloher RE, Tomana M: Deficient galactosylation of serum IgG in inflammatory bowel disease: correlation with disease activity. *J Clin Gastroenterol* 18: 86–87, 1994
14. Trbojević-Akmačić I, Ventham NT, Theodoratou E, Vuckovic F: kennedy, NA, Kristic, J, Nimmo, ER, Drummond, D, Stambuk, J, Klaric, L, Dunlop, MG, Novokmet, M, Aulchenko, Y, Gornik, O, Kolarich, D, Wuhrer, M, McGovern, D, Annese, V, Kalla, R, Pemberton, JM, Spencer, D, Zoldos, V, Fernandes, D, Campbell, H, Pucic Bakovic, M, Satsangi, J, Lauc, G: Inflammatory bowel disease associates with pro-inflammatory potential of the IgG glycome. *Inflamm Bowel Dis* 21(6): 1237–1247, 2015
15. Ohtsubo K, Marth JD: Glycosylation in cellular mechanisms of health and disease. *Cell* 126: 855–867, 2006
16. Gornik O, Lauc G: Glycosylation of serum proteins in inflammatory diseases. *Dis Markers* 25: 267–278, 2008
17. Moore JS, Wu X, Kulhavy R, Tomana M, Novak J, Moldoveanu Z, Brown R, Goepfert PA, Mestecky J: Increased levels of galactose-deficient IgG in sera of HIV-1-infected individuals. *AIDS* 19: 381–389, 2005
18. Paragh G, Seres I, Harangi M, Fülöp P: Dynamic interplay between metabolic syndrome and immunity. *Adv Exp Med Biol* 824: 171–190, 2014
19. Camps J, García-Heredia A: Introduction: oxidation and inflammation, a molecular link between non-communicable diseases. *Adv Exp Med Biol* 824: 1–4, 2014
20. Lopez-Parra V, Mallavia B, Lopez-Franco O, Ortiz-Muñoz G, Oguiza A, Recio C, Blanco J, Nimmerjahn F, Egido J, Gomez-Guerrero C: Fcγ receptor deficiency attenuates diabetic nephropathy. *J Am Soc Nephrol* 23: 1518–1527, 2012
21. van Timmeren MM, van der Veen BS, Stegeman CA, Petersen AH, Hellmark T, Collin M, Heeringa P: IgG glycan hydrolysis attenuates ANCA-mediated glomerulonephritis. *J Am Soc Nephrol* 21: 1103–1114, 2010
22. Novak J, Julian BA, Mestecky J, Renfrow MB: Glycosylation of IgA1 and pathogenesis of IgA nephropathy. *Semin Immunopathol* 34: 365–382, 2012
23. Novak J, Tomana M, Matousovic K, Brown R, Hall S, Novak L, Julian BA, Wyatt RJ, Mestecky J: IgA1-containing immune complexes in IgA nephropathy differentially affect proliferation of mesangial cells. *Kidney Int* 67: 504–513, 2005
24. Menni C, Keser T, Mangino M, Bell JT, Erte I, Akmačić I, Vučković F, Pučić Baković M, Gornik O, McCarthy MI, Zoldoš V, Spector TD, Lauc G, Valdes AM: Glycosylation of immunoglobulin g: role of genetic and epigenetic influences. *PLoS ONE* 8: e82558, 2013
25. Pivac N, Knezevic A, Gornik O, Pucic M, Igl W, Peeters H, Crepel A, Steyaert J, Novokmet M, Redzic I, Nikolac M, Hercigonja VN, Curkovic KD, Curkovic M, Nedic G, Muck-Seler D, Borovecki F, Rudan I, Lauc G: Human plasma glycome in attention-deficit hyperactivity disorder and autism spectrum disorders. *Mol Cell Proteomics* 10: M110 004200, 2011.
26. Novokmet M, Lukić E, Vučković F, Đurić Ž, Keser T, Rajšl K, Remondini D, Castellani G, Gašparović H, Gornik O, Lauc G: Changes in IgG and total plasma protein glycomes in acute systemic inflammation. *Sci Rep* 4: 4347, 2014
27. Parekh RB, Dwek RA, Sutton BJ, Fernandes DL, Leung A, Stanworth D, Rademacher TW, Mizuochi T, Taniguchi T, Matsuta K, Takeuchi F, Nagano Y, Miyamoto T, Kobata A: Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature* 316: 452–457, 1985
28. Krištić J, Vučković F, Menni C, Klarić L, Keser T, Beceheli I, Pučić-Baković M, Novokmet M, Mangino M, Thaqi K, Rudan P, Novokmet N, Sarac J, Missoni S, Kolčić I, Polašek O, Rudan I, Campbell H, Hayward C, Aulchenko Y, Valdes A, Wilson JF, Gornik O, Primorac D, Zoldoš V, Spector T, Lauc G: Glycans are a novel biomarker of chronological and biological ages. *J Gerontol A Biol Sci Med Sci* 69: 779–789, 2014
29. Mihai S, Nimmerjahn F: The role of Fc receptors and complement in autoimmunity. *Autoimmun Rev* 12: 657–660, 2013
30. Malhotra R, Wormald MR, Rudd PM, Fischer PB, Dwek RA, Sim RB: Glycosylation changes of IgG associated with rheumatoid arthritis can activate complement via the mannose-binding protein. *Nat Med* 1: 237–243, 1995
31. Cook HT: Complement and kidney disease. *Curr Opin Nephrol Hypertens* 22: 295–301, 2013
32. Rombouts Y, Ewing E, van de Stadt LA, Selman MH, Trouw LA, Deelder AM, Huizinga TW, Wuhrer M, van Schaardenburg D, Toes RE, Scherer HU: Anti-citrullinated protein antibodies acquire a pro-inflammatory Fc glycosylation phenotype prior to the onset of rheumatoid arthritis. *Ann Rheum Dis* 74: 234–241, 2015
33. Ercan A, Cui J, Chatterton DE, Deane KD, Hazen MM, Brintnell W, O'Donnell CI, Derber LA, Weinblatt ME, Shadick NA, Bell DA, Cairns E, Solomon DH, Holers VM, Rudd PM, Lee DM: Aberrant IgG galactosylation precedes disease onset, correlates with disease activity, and is prevalent in autoantibodies in rheumatoid arthritis. *Arthritis Rheum* 62: 2239–2248, 2010
34. Tomana M, Schrohenloher RE, Koopman WJ, Alarcón GS, Paul WA: Abnormal glycosylation of serum IgG from patients with chronic inflammatory diseases. *Arthritis Rheum* 31: 333–338, 1988
35. Tomana M, Schrohenloher RE, Reveille JD, Arnett FC, Koopman WJ: Abnormal galactosylation of serum IgG in patients with systemic lupus erythematosus and members of families with high frequency of autoimmune diseases. *Rheumatol Int* 12: 191–194, 1992
36. Anthony RM, Ravetch JV: A novel role for the IgG Fc glycan: the anti-inflammatory activity of sialylated IgG Fcs. *J Clin Immunol* 30[Suppl 1]: S9–S14, 2010
37. Kaneko Y, Nimmerjahn F, Ravetch JV: Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. *Science* 313: 670–673, 2006
38. Anthony RM, Nimmerjahn F, Ashline DJ, Reinhold VN, Paulson JC, Ravetch JV: Recapitulation of IVIG anti-inflammatory activity with a recombinant IgG Fc. *Science* 320: 373–376, 2008
39. Huffman JE, Pučić-Baković M, Klarić L, Hennig R, Selman MH, Vučković F, Novokmet M, Krištić J, Borowiak M, Muth T, Polašek O, Razdorov G,

Gornik O, Plomp R, Theodoratou E, Wright AF, Rudan I, Hayward C, Campbell H, Deelder AM, Reichl U, Aulchenko YS, Rapp E, Wuhrer M, Lauc G: Comparative performance of four methods for high-throughput glycosylation analysis of immunoglobulin G in genetic and epidemiological research. *Mol Cell Proteomics* 13: 1598–1610, 2014

40. Karsten CM, Pandey MK, Figge J, Kilchenstein R, Taylor PR, Rosas M, McDonald JU, Orr SJ, Berger M, Petzold D, Blanchard V, Winkler A, Hess C, Reid DM, Majoul IV, Strait RT, Harris NL, Köhl G, Wex E, Ludwig R, Zillikens D, Nimmerjahn F, Finkelman FD, Brown GD, Ehlers M, Köhl J: Anti-inflammatory activity of IgG1 mediated by Fc galactosylation and association of FcγRIIB and dectin-1. *Nat Med* 18: 1401–1406, 2012
41. Ferrara C, Stuart F, Sondermann P, Brünker P, Umaña P: The carbohydrate at FcgammaRIIIa Asn-162. An element required for high affinity binding to non-fucosylated IgG glycoforms. *J Biol Chem* 281: 5032–5036, 2006
42. Pucic M, Knezevic A, Vidic J, Adamczyk B, Novokmet M, Polasek O, Gornik O, Supraha-Goreta S, Wormald MR, Redzic I, Campbell H, Wright A, Hastie ND, Wilson JF, Rudan I, Wuhrer M, Rudd PM, Josic D, Lauc G: 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, 2011.
43. Masuda K, Kubota T, Kaneko E, Iida S, Wakitani M, Kobayashi-Natsume Y, Kubota A, Shitara K, Nakamura K: Enhanced binding affinity for FcgammaRIIIa of fucose-negative antibody is sufficient to induce maximal antibody-dependent cellular cytotoxicity. *Mol Immunol* 44: 3122–3131, 2007
44. Moayyeri A, Hammond CJ, Valdes AM, Spector TD: Cohort Profile: TwinsUK and healthy ageing twin study. *Int J Epidemiol* 42: 76–85, 2013
45. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF 3rd, Feldman HI, Kusek JW, Eggers P, Van Lente F, Greene T, Coresh J; CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration): A new equation to estimate glomerular filtration rate. *Ann Intern Med* 150: 604–612, 2009
46. Stevens PE, Levin A; Kidney Disease: Improving Global Outcomes Chronic Kidney Disease Guideline Development Work Group Members: Evaluation and management of chronic kidney disease: synopsis of the kidney disease: improving global outcomes 2012 clinical practice guideline. *Ann Intern Med* 158: 825–830, 2013

This article contains supplemental material online at http://jasn.asnjournals.org/lookup/suppl/doi:10.1681/ASN.2015010109/-/DCSupplemental.

4. The Association Between Low Back Pain and Composition of IgG Glycome



OPEN

The Association Between Low Back Pain and Composition of IgG Glycome

Received: 08 December 2015
Accepted: 09 May 2016
Published: 27 May 2016

Maxim B. Freidin[1,*], Toma Keser[2,*], Ivan Gudelj[3], Jerko Štambuk[3], Dunja Vučenović[1,4], Massimo Allegri[5], Tamara Pavić[2], Mirna Šimurina[2], Stella M. Fabiane[1], Gordan Lauc[2,3,#] & Frances M. K. Williams[1,#]

Low back pain (LBP) is a common debilitating condition which aetiology and pathogenesis are poorly understood. We carried out a first so far analysis of associations between LBP and plasma IgG N-glycome in a sample of 4511 twins from TwinsUK database assessed for LBP, lumbar disc degeneration (LDD) as its possible cause, and IgG-glycan levels. Using weighted correlation network analysis, we established a correlation between LBP and glycan modules featured by glycans that either promote or block antibody-dependent cell-mediated cytotoxicity (ADCC). The levels of four glycan traits representing two of those modules were statistically significantly different in monozygotic twins discordant for LBP. Also, the trend to higher prevalence of systemic inflammatory disorders was shown for twins with low level of fucosylated glycans and high level of non-fucosylated glycans. Core fucosylation of IgG is a "safety switch" reducing ADCC, thus our results suggest the involvement of ADCC and associated inflammation in pathogenesis of LBP. No correlation between LDD scores and glycans was found assuming that the inflammation may not be a part of LDD. These data provide a new insight into understanding the complex pathophysiology of LBP and suggest glycan levels as a possible biomarker for inflammation-related subtypes of LBP.

Low back pain (LBP) is a common musculoskeletal condition in all ages[1]. The lifetime prevalence of non-specific LBP may reach 80%, with the annual prevalence ranging between 25% and 60% in different ethnic groups[2,3]. It is a diverse group of mixed pain syndromes with different molecular pathologies at different structural levels displaying similar clinical manifestations and radiologic findings. Why there is such huge inter-personal variability in severity of chronic LBP is yet to be clearly defined. Lumbar disk degeneration (LDD) is widely believed to be one of the major contributing factors. Nevertheless, MRI findings of disc degeneration cannot help to define clearly the pathophysiology of LBP and its prognosis[4].

Even though, the development of LDD is associated with such occupational factors as heavy lifting, frequent bending and twisting[5], genetic predisposition is much more important as a risk factor[6].

A TwinsUK study showed genetic background as the major factor associated with LBP in women and also revealed a significant genetic correlation between LBP and LDD[7]. Genetic studies identified a dozen of genes associated with LDD, such as genes coding collagens, vitamin D receptor, interleukins, matrix metalloproteinases and other molecules[8,9].

Large genome-wide linkage study and a genome-wide meta-analysis identified *CHST3* gene associated with LDD, and a subsequent functional analysis showed the risk allele decreases the gene expression, possibly, due to the enhanced interaction with miR-513a-5p microRNA[10]. Also, a meta-analysis of several genome-wide association studies revealed an association between LDD and *PARK* gene with the differential methylation of the *PARK* gene promoter as a possible cause for the association[11]. Also, an increased methylation of *SPARC* gene was found

[1]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [2]Department of Biochemistry and Molecular Biology, Faculty of Pharmacy and Biochemistry, University of Zagreb, Zagreb, Croatia. [3]Genos Glycoscience Research Laboratory, Zagreb, Croatia. [4]Department of Molecular Biology, Faculty of Science, University of Zagreb, Zagreb, Croatia. [5]Department of Surgical Science, University of Parma; Anaesthesia Intensive Care and Pain Therapy Service, Parma Hospital, Parma, Italy. *These authors contributed equally to this work. #These authors jointly supervised this work. Correspondence and requests for materials should be addressed to G.L. (email: glauc@pharma.hr) or F.M.K.W. (email: frances.williams@kcl.ac.uk)

Trait	Value	%
Age ± SD	51.8 ± 14.1	–
Females/males	4175/336	92.6/7.4
BMI ± SD	26.3 ± 5.0	–
LBP in total sample, positive/negative/unknown	1064/2493/954	23.6/55.3/21.1
LBP in pairs of twins, number of pairs		
MZ twins		
both positive/both negative/discordant/unknown in at least one of the twins	57/252/126/56	11.6/54.7/25.7/11.4
DZ twins		
both positive/both negative/discordant/unknown in at least one of the twins	146/452/316/301	12.0/37.2/26.0/24.8

Table 1. Demographics of the studied twins.

to be associated with LBP and LDD in humans and mice[12]. These studies incur epigenetic factors in the development of LDD and LBP.

Thus, the discovery of molecular factors contributing to the predisposition to LBP and LDD and mechanisms by which these factors act is essential to facilitate the development of new biomarkers of risk and response to specific treatments. Apart from genomic and epigenomic factors, other newly established "omes" can be of value.

In particular, glycome (the entire composition of glycans) attracts attention. Glycans constitute the most abundant and diverse form of the post-translational modifications. All cell surface and secreted glycoproteins that contain appropriate sequences (Asn-X-Ser/Thr where X is any amino acid except proline) can potentially acquire N-linked oligosaccharides (N-glycans) while they travel through the endoplasmic reticulum and the Golgi compartments. Glycans can influence disease development in many syndromes such as congenital disorders of glycosylation, cancer, rheumatoid arthritis and AIDS[13]. Glycans are crucial for the immune system, as some of the most important interactions between the immune system and viruses and bacteria are mediated by protein-glycan interactions. The biological functions of glycans go from basic structural roles to development, protein folding and immune response.

While genes unequivocally determine the structure of each polypeptide, there is no genetic template for the glycan part. Instead, hundreds of genes and their products interact in the complex pathway of glycan biosynthesis resulting in a very complex biosynthetic pathway that is further complicated by both direct environmental influence (nutrition, hormonal status, etc) and epigenetic memory of past environmental effects (altered gene expression)[14–16]. It is possible that some of the considerable genetic predisposition to LDD may be mediated via glycans.

Immunoglobulin G (IgG) glycosylation has been well defined, and many important functional effects of alternative IgG glycosylation have been described[17]. Glycans that lack terminal galactose activate complement and make IgG pro-inflammatory, while the addition of galactose decreases inflammatory potential of IgG[18,19]. Further extension of IgG glycans by the addition of sialic acid dramatically changes the physiological role of IgG, converting it from a pro-inflammatory into an anti-inflammatory agent. Terminal α2,6-sialylation of IgG glycans decreases the ability of IgG to bind to activating Fcγ Rs and promotes recognition by DC-SIGN, which increases expression of inhibitory FcγRIIB and is anti-inflammatory[20]. Another example is the role of core fucose in the modulation of antibody-dependent cellular cytotoxicity: IgG-containing glycans that lack core fucose have 100-fold increased affinity for FcγRIIIA and are therefore much more efficient in activating antibody-dependent cellular cytotoxicity than fucosylated glycoforms of the same molecule[21].

In this study, we analyzed twins from TwinsUK registry, to assess whether persons reporting episodes of LBP had detectable levels of altered IgG glycosylation.

Results

The study aimed at identifying relationships between IgG glycosylation and pain. We used a cohort of twins from TwinsUK registry with established phenotypes of LBP and using a recently developed high-throughput analysis method quantified IgG glycans in their plasma specimens. After pre-processing and filtering of the data, a total of 4511 individuals were analyzed including 1215 pairs of DZ twins, 491 pairs of MZ twins, and 1099 unpaired individuals (Table 1). Low back pain status was known for 3557 individuals.

Analysis of association between glycan levels and LBP. Overall, 76 directly measured or derived glycan traits were assessed for an association with LBP (Supplementary Table 1). Linear mixed model analysis with BMI, sex, inflammatory diseases and LBP status included as fixed covariate and variation in IgG glycan quantities within twin pairs as random effect revealed nominally statistically significant associations of LBP with several glycan traits with the strongest association seen for IGP49 ($GP10^n$) (Fig. 1). However, none of the associations passed the significance threshold set to control for multiple testing ($p = 0.0027$). Same was true for the analysis of correlations between glycan levels and summary scores for lumbar magnetic resonance imaging signs (LSUM; Supplementary Table 2; Supplementary Figure 1).

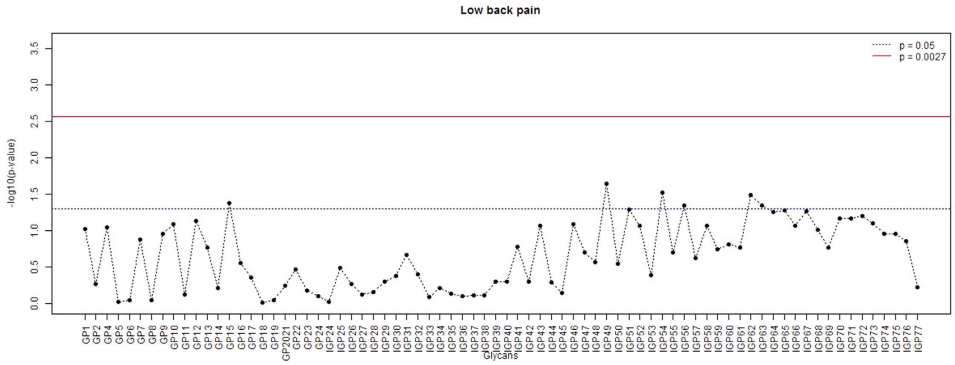
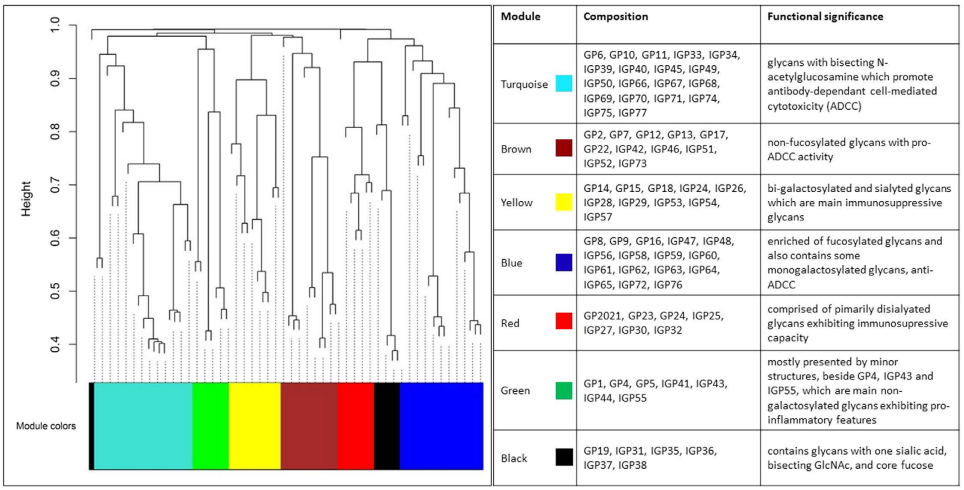


Figure 1. P-values (−log10) for the analysis of associations between glycan levels and low back pain. Linear mixed models were used to estimate the associations using LBP status, BMI, sex, and major inflammatory disease status as fixed factors and family status as a random factor.



Module	Composition	Functional significance
Turquoise	GP6, GP10, GP11, IGP33, IGP34, IGP39, IGP40, IGP45, IGP49, IGP50, IGP66, IGP67, IGP68, IGP69, IGP70, IGP71, IGP74, IGP75, IGP77	glycans with bisecting N-acetylglucosamine which promote antibody-dependant cell-mediated cytotoxicity (ADCC)
Brown	GP2, GP7, GP12, GP13, GP17, GP22, IGP42, IGP46, IGP51, IGP52, IGP73	non-fucosylated glycans with pro-ADCC activity
Yellow	GP14, GP15, GP18, IGP24, IGP26, IGP28, IGP29, IGP53, IGP54, IGP57	bi-galactosylated and sialyted glycans which are main immunosuppressive glycans
Blue	GP8, GP9, GP16, IGP47, IGP48, IGP56, IGP58, IGP59, IGP60, IGP61, IGP62, IGP63, IGP64, IGP65, IGP72, IGP76	enriched of fucosylated glycans and also contains some monogalactosylated glycans, anti-ADCC
Red	GP2021, GP23, GP24, IGP25, IGP27, IGP30, IGP32	comprised of pimarily disialyated glycans exhibiting immunosupressive capacity
Green	GP1, GP4, GP5, IGP41, IGP43, IGP44, IGP55	mostly presented by minor structures, beside GP4, IGP43 and IGP55, which are main non-galactosylated glycans exhibiting pro-inflammatory features
Black	GP19, IGP31, IGP35, IGP36, IGP37, IGP38	contains glycans with one sialic acid, bisecting GlcNAc, and core fucose

Figure 2. Modules of correlated glycans obtained using WGCNA methodology.

WGCNA. Using the weighted correlation network analysis (WGCNA) methodology, we carried out a network analysis for glycan levels to establish clusters of correlated glycans which, possibly, reflect their functional relationships and revealed associations between these clusters and pain phenotypes.

Using signed networks, we identified seven modules of correlated glycans (Fig. 2; Supplementary Table 1), which can be grouped into two big branches comprising yellow, brown and turquoise modules, from one hand, and black, green, blue and red modules, from the other hand (Fig. 3).

The most abundand turquoise module is comprised of glycans with bisecting *N*-acetylglucosamine (GlcNAc) which was reported to promote antibody-dependant cell-mediated cytotoxicity (ADCC)[22]. Similarly, the brown module belonging to the same branch of modules, contains non-fucosylated glycans, which also promote ADCC. The yellow module from the same branch is enriched of bi-galactosylated and sialylated glycans which are main immunosuppressive glycans.

Another branch's biggest blue module, in opposition to the turquoise and brown modules, is enriched of fucosylated glycans, does not include glycans with bisecting GlcNAc, and also contains some monogalactosylated glycans. The red module is comprised of primarily disialyated glycans exhibiting immunosupressive capacity. The green module is mostly presented by minor structures, beside GP4, IGP43 ($GP4^n$) and IGP55 ($G0^n$), which are main non-galactosylated glycans exhibiting pro-inflammatory features. Finally, the black module contains glycans with one sialic acid, bisecting GlcNAc, and core fucose.

To reveal relationships between glycan modules and pain phenotypes, we carried out a correlation analysis between module eigenvalues (estimated as first principal component of glycan levels in a module) and LBP and MRI trait LSUM (Fig. 4). LBP was found to be positively correlated with turquoise and brown modules and negatively with blue module. A hint to correlation between green module and LBP was seen; however, these correlations did not reach statistical significance.

Also, even though the correlation strength between glycan modules and MRI-traits was of similar magnitude as those for LBP (R = 0.04–0.05), the correlations did not reach statistical significance for MRI-traits.

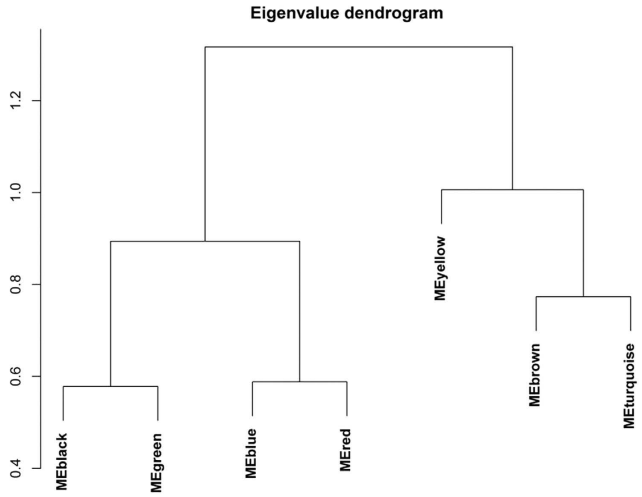


Figure 3. Relationships between modules of correlated glycans.

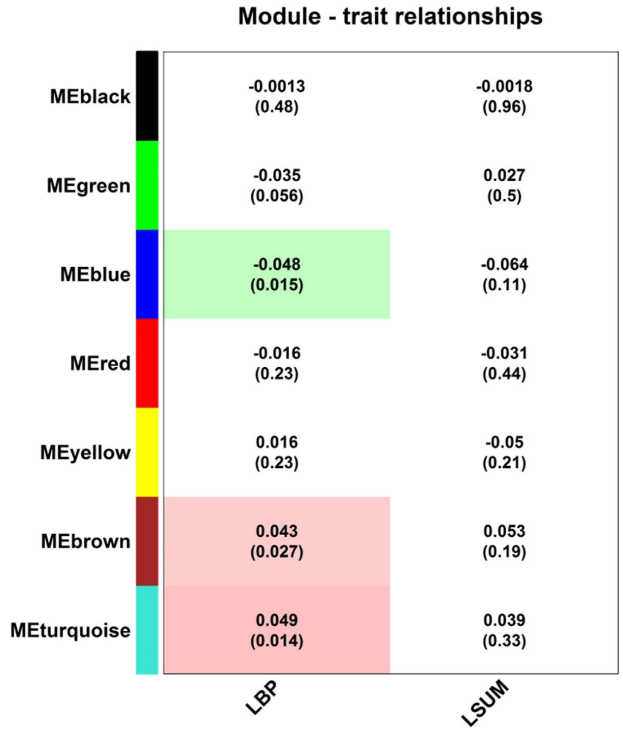


Figure 4. Correlations between module eigenvalues and pain phenotypes. Correlations were calculated between module eigenvalues (vector of first principal component of glycans in a module) and low back pain (LBP) using point-biserial correlation coefficient and summary score for magnetic resonance imaging signs for lumbar spine (LSUM) using Pearson correlation coefficients. Corresponding p-values are provided in brackets.

Average glycan significance for this phenotype (defined as the average for the correlation coefficients between glycan levels in a module and a trait) was highest for blue and turquoise modules for LBP (Fig. 5). These results suggest that glycans from the blue and turquoise modules may be of especial interest for subsequent study of their relationships with pain phenotype.

Discordant twins analysis. To further analyse the relationships between glycome and pain phenotypes, we carried out comparisons of glycan levels in MZ and DZ twins discordant for LBP using paired t-test.

For MZ twins, we identified statistically significant differences between the twins with and without LBP for the IGP65 ($FG2^n/G2^n$), IGP74 ($FBG2^n/FG2^n$), IGP75 ($FBG2^n/[FG2^n + FBG2^n]$), and IGP76 ($FG2^n/[BG2^n + FBG2^n]$)

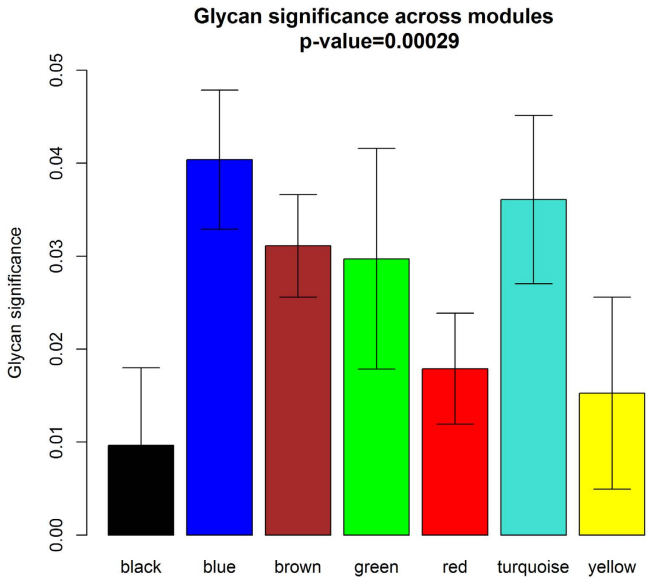


Figure 5. Average glycan significance across modules for LBP. Glycan significance was defined as the average coefficient of correlation between a trait and glycan levels in a module; p-value is given for Kruskal-Wallis test for the difference of glycan significance across the modules.

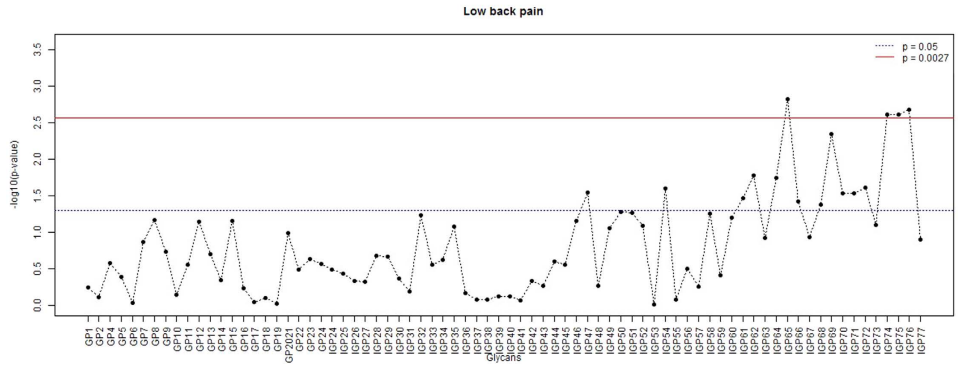


Figure 6. P-values (−log10) for comparisons of mean glycan levels in MZ twins discordant for LBP phenotype by paired t-test. Red line corresponds to p = 0.0027 which was taken as the significance threshold based on the 19 effective independent tests with Sidak's correction for multiple testing.

derived traits (Fig. 6; $p < 0.0027$). Notably, these four glycan traits belong to the blue and turquoise modules identified in the WGCNA analysis. Accordingly, IGP65 and IGP76 of the blue module were found to be elevated in MZ twins without LBP, while IGP74 and IGP75 of the turquoise module were elevated in MZ twins with LBP (Fig. 7). The four glycan traits were derived from neutral glycans GP14 and GP15, and also GP13 for IGP76, with GP14 being the numerator for IGP65 and IGP76, while GP15 the numerator for the other two (Supplementary Table 1). Intriguingly, neither GP14, nor GP15 showed any trend to association with LBP; however, there was a weak, but significant negative correlation between GP14 and LCUM values (Pearson $r = -0.08$, $p = 0.04$; Supplementary Table 2; Supplementary Figure 1).

No statistical significant differences in glycan levels were found for DZ twins or MZ and DZ twins combined.

To pursue a cause for association between LBP and glycan levels in MZ twins discordant for LBP, we split them into groups of high and low level of IGP65, IGP74, IGP75, and IGP76 using 25% and 75% quintiles as the cut off points and compared the prevalence of systemic inflammatory disorders (rheumatoid arthritis, systemic lupus erythematosus, ulcerative colitis and Crohn's disease) in these groups. We found the increase of inflammatory diseases in individuals exhibiting low levels of IGP65 and IGP76 and high levels of IGP74 and IGP75 (Fig. 8). This pattern was in full agreement with the observation of association between these glycan levels and LBP, though the differences in inflammatory disorders prevalence did not reach statistical significance (according to Fisher's exact test p-values).

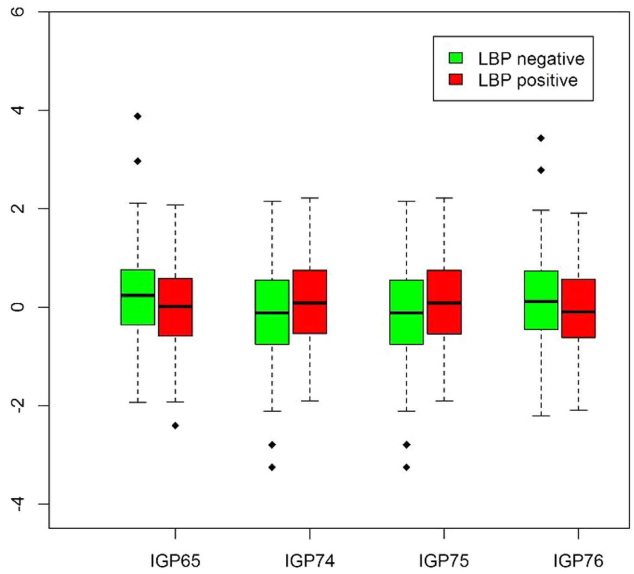


Figure 7. Glycan levels in MZ twins discordant for LBP phenotype.

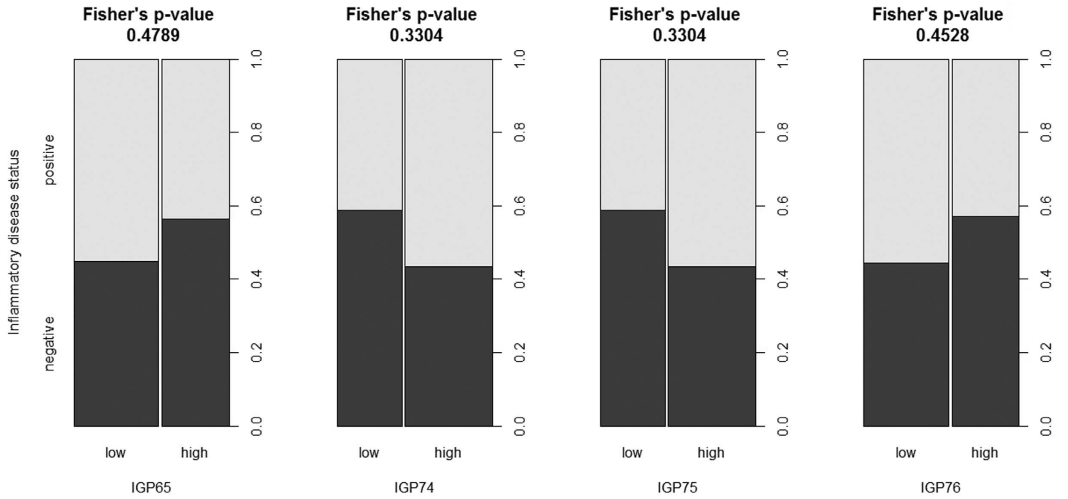


Figure 8. The prevalence of systemic inflammatory disorders (rheumatoid arthritis, systemic lupus erythematosus, ulcerative colitis, and Crohn's disease) in twins with high and low levels of glycans and discordant for LBP. The cut off points for glycans levels were set at 25% and 75% quintile for the corresponding distribution.

Discussion

In this study we have evaluated association between levels of plasma IgG glycans and LBP. Linear mixed-models analysis did not reveal statistically significant ($p < 0.0027$) associations between glycan levels and LBP. However, for several glycans nominally significant associations were obtained.

In an attempt to consider glycome as a whole, we carried out a network analysis using weighted correlation network approach (WGCNA). This is a powerful methodology for revealing clusters (modules) of multiple omic traits, such as genome-wide gene expression or global methylation profiles, and placing them into a biological context through the analysis of associations between the clusters and diseases or traits of interest[23–28]. To the best of our knowledge, this method has never been applied before to glycome. Even though, glycans do not interact with each other in a way of genes or proteins, the network methodology underlying WGCNA analysis still seems valuable for glycome as it allows revealing functionally related groups of glycans exhibiting overlapping biological activity.

Using WGCNA approach we revealed seven modules of glycans clustered according to their functional capabilities, with the two biggest modules (turquoise and blue) enriched with glycans with opposite potential for the development of ADCC through the regulation of core fucosylation and bisection. Fucosylation is crucial in many biological processes and inflammation in particular. On average, 95% of the IgG population is core

fucosylated[29]; core fucose prevents activation of ADCC, thus, most of the immunoglobulins have a "safety switch", which prevents them from killing the target cell. Malfunction of this system appears to be associated with autoimmune diseases, as indicated by both pleiotropic effects of genes that associate with IgG glycosylation on different inflammatory and autoimmune diseases, and observed alterations in IgG glycosylation in systemic lupus erythematous[30,31] and many inflammatory diseases[32]. We observed a positive correlation between LBP and "pro-ADCC" turquoise and brown modules and a negative correlation between LBP and "anti-ADCC" blue module (Fig. 4). Assuming that the development of LBP syndrome is in part related to inflammation[33] and that ADCC contributes to the joint inflammation in some types of back pain[34], one could expect that the decreased levels of core fucosylation and increased levels of bisecting GlcNAc in IgG glycans may contribute to increased ADCC and inflammation in LBP patiens. This observation corroborates with the finding of the significant difference in the levels of blue and turquoise module glycans in MZ twins discordant for LBP: IGP65 and IGP76 (blue module) were decreased in LBP-positive persons, while IGP74 and IGP75 (turquoise module) were increased in LBP-positive persons (Figs 6 and 7). Accordingly, we found the increased prevalence of major inflammatory diseases in LBP-discordant MZ twins with low levels of IGP65 and IGP76 and high levels of IGP75 and IGP76 (Fig. 8). Even though, the differences in the diseases prevalence were not statistically significant, the pattern of the differences corresponds to the pattern of association between the glycan levels and LBP, thus linking the three entities (glycans, LBP, and systemic inflammatory diseases). It is worth noting, that during inflammation process IgG glycome may vary in a quite complex way following several different patterns[35]. Therefore, the relationships between fucosylation, bisecting GlcNAc and LBP may not be entirely straightforward.

Interestingly, no statistically significant differences were found between DZ twins discordant for LBP. This may reflect a pronounced impact of environmental or gene-environment variability on co-variation between glycan levels and LBP.

Conclusion

The current study was a first attempt to establish relationships between LBP and glycome. We proceeded from a hypothesis that LBP may be associated with an occult inflammation reflected by IgG glycan levels. We found consistent associations between pro- and anti-ADCC glycans with LBP, thus providing a proof for the tested hypothesis. Overall, our findings provide a further clue how inter-individual differences in IgG glycosylation might affect mechanisms of the development of LBP and suggest that glycans can be of interest as possible patient stratification biomarkers of this pain syndrome.

Material and Methods

Sample. Participants were a sample of MZ and DZ twins enlisted in the Twins UK registry[36]. The participants in the present study had undergone height and weight measurements used to calculate BMI. Collection of socio-demographic and LBP data was carried out during clinical visit or via a postal self-completion questionnaire. The twins were unaware of the precise research hypothesis addressed in the present study.

The study was carried out under the auspices of the FP7 PainOmics project and was approved by the St Thomas' Hospital Research Ethics Committee. All the methods were carried out in accordance with the approved guidelines. All participating twins provided signed informed consent.

Participating twins underwent an assessment that included a nurse-led interview and a number of clinical and laboratory tests. As part of the study, the twins completed two standardized questionnaires relating to their lifetime history of low back symptoms. The questionnaires have been completed by each twin separately. The questionnaires included written questions and a mannequin pain diagram allowing an assessment of the timing, distribution, radiation, severity, and duration of pain together with information relating to functional disability. Low back pain was defined on a mannequin as being located between the 12th rib and the gluteal folds. First questionnaire followed the format of questions used in the UK Medical Research Council Nurses Study[37] and the procedure of the assessment is detailed elsewhere[38]. The assessment of twins using this questionnaire was done in framework of the UK Twin Spine Study[7,39]. Another questionnaire followed the format of London Fibromyalgia Epidemiology Symptom Screening Questionnaire[40]. Specifically, the following questions have been used: "In the past month, have you had pain symptoms in this area (central lumbar region, left lumbar region, right lumbar region, left buttock, right buttock) lasting at least 24 hours? " and "Have you had pain like this in this areas for at least the past 3 months? ". This assessment using this questionnaire was done in framework of ongoing studies of chronic pain syndromes undertaken by the Department of Twin Research and Genetic Epidemiology at King's College London[41].

The LBP phenotype was defined as a binary trait based on questionnaire responses (1 = affected and 0 = non-affected). Participants were categorised as cases for LBP if they reported having the syndrome with a total duration of >1 month and associated with disability according to the first questionnaire or at least 3 months according to the second questionnaire. Overall, 1656 and 2975 partly overlapping participants have been assessed using first and second questionnaire, respectively, which allowed identification of the LBP status in 3557 participants. Out of them, 585 participants (35.3% tested using first questionnaire) had disabling LBP lasting >1 month and 582 participants (20.0% tested using second questionnaire) had LBP lasting at least 3 months.

For 647 participants magnetic resonance imaging (MRI) was carried out as a part of LBP status assessment. The MRI scan was performed using a Siemens (Munich, Germany) 1.0 T superconducting magnet. Sagittal images were obtained using a fast spin-echo sequence of time to recovery (TR)/time to echo (TE) 5000–4500/112 msec, with a slice thickness of 4 mm. Grading was performed on T2-weighted images, although T1 images were also obtained for certain measurements. Axial sections were obtained at selected levels to assess structural changes in individuals who had features suggesting prolapse. To avoid problems related to diurnal variation in disc height all MRI scans were performed >1 hour after the subjects arose from sleep in the morning, with no exercise or other rest allowed between arising and the scan, and importantly, each twin pair was scanned at the same appointment

and on the same machine[42]. A disease (LBP) severity score was constructed from the sum of scores for disc bulge, height, signal change, and narrowing in the lumbar spine (LSUM).

Analysis of IgG glycans

Isolation of IgG from Human Plasma. The IgG was isolated using protein G monolithic plates (BIA Separations, Ajdovščina, Slovenia) as described previously[29]. Briefly, 50 to 90 μL of serum was diluted 7 × with 1 × PBS, pH 7.4, applied to the protein G plate and instantly washed with 1 × PBS, pH 7.4, to remove unbound proteins. IgG was eluted with 1 mL of 0.1 M formic acid (Merck, Darmstadt, Germany) and neutralized with 1 M ammonium bicarbonate (Merck).

Glycan Release and Labeling. IgG samples were first denatured with addition of 30 μL 1.33% sodium dodecyl sulfate (w/v) (Invitrogen, Carlsbad, CA) and by incubation at 65 °C for 10 min. Subsequently, 10 μL of 4% Igepal-CA630 (Sigma-Aldrich, St. Louis, MO) and 1.25 mU of PNGase F (ProZyme) in 10 μL 5 × phosphate-buffered saline were added to the samples. The samples were incubated overnight at 37 °C for N-glycan release. The released N-glycans were labeled with 2-AB. The labeling mixture was freshly prepared by dissolving 2-AB (Sigma–Aldrich) in dimethyl sulfoxide (Sigma–Aldrich) and glacial acetic acid (Merck) mixture (85:15, v/v) to a final concentration of 48 mg/mL. A volume of 25 μL of labeling mixture was added to each N-glycan sample in the 96-well plate. Also, 25 μL of freshly prepared reducing agent solution (106.96 mg/mL 2-picoline borane [Sigma-Aldrich] in dimethyl sulfoxide) was added and the plate was sealed using adhesive tape. Mixing was achieved by shaking for 10 min, followed by 2-hour incubation at 65 °C. Samples (in a volume of 100 μL) were brought to 80% acetonitrile (ACN) (v/v) by adding 400 μL of ACN (J.T. Baker, Phillipsburg, NJ). Free label and reducing agent were removed from the samples using hydrophilic interaction chromatography–solid-phase extraction. An amount of 200 μL of 0.1 g/mL suspension of microcrystalline cellulose (Merck) in water was applied to each well of a 0.45 μm GHP filter plate (Pall Corporation, Ann Arbor, MI). Solvent was removed by application of vacuum using a vacuum manifold (Millipore Corporation, Billerica, MA). All wells were prewashed using 5 × 200 μL water, followed by equilibration using 3 × 200 μL acetonitrile/water (80:20, v/v). The samples were loaded to the wells. The wells were subsequently washed seven times using 200 μL acetonitrile/water (80:20, v/v). Glycans were eluted two times with 100 μL of water and combined eluates were stored at −20 °C until usage.

Hydrophilic Interaction Chromatography (HILIC)–Ultra Performance Liquid Chromatography. Fluorescently labeled N-glycans were separated by hydrophilic interaction chromatography on a Waters Acquity ultra performance liquid chromatography (UPLC) instrument (Milford, MA) consisting of a quaternary solvent manager, sample manager, and an FLR fluorescence detector set with excitation and emission wavelengths of 250 and 428 nm, respectively. The instrument was under the control of Empower 2 software, build 2145 (Waters). Labeled N-glycans were separated on a Waters bridged ethylene hybrid, glycan chromatography column, 100 × 2.1 mm internal diameter, 1.7-μm bridged ethylene hybrid particles, with 100 mM ammonium formate, pH 4.4, as solvent A and acetonitrile as solvent B. The separation method used a linear gradient of 75% to 62% acetonitrile (vol/vol) at flow rate of 0.4 mL/min in a 25-minute analytical run. Samples were maintained at 5 °C before injection, and the separation temperature was 60 °C. The system was calibrated using an external standard of hydrolyzed and 2-AB labeled glucose oligomers from which the retention times for the individual glycans were converted to glucose units. Data processing was performed using an automatic processing method with a traditional integration algorithm after which each chromatogram was manually corrected to maintain the same intervals of integration for all the samples. The chromatograms were all separated in the same manner into 24 peaks (GP1-GP24).

Statistical analysis. *Pre-processing and filtering.* Directly measured glycan levels were normalized and experimental noise was removed through filtering and batch correction. Before this, we removed GP3 and combined GP20 and GP21 into a single trait.

First, we filtered out most extreme values from the dataset (beyond 0.999% percentile). Then, quotient normalization was applied using median values across the dataset as a reference[43]. Batch effect associated with different plates used to measure glycan levels was identified and corrected for using ratio-based method with either geometric mean or median[44]. As the results were almost equivalent, herewith, we report only the results for the dataset corrected with geometric mean.

After these steps, we estimated 55 derived glycan levels from the directly measured glycans[45] using *glycanr* package for R [https://github.com/iugrina/glycanr] (Supplementary Table 1). These derived traits average particular glycosylation features (galactosylation, fucosylation, sialylation) across different individual glycan structures and consequently they are more closely related to individual enzymatic activities and underlying genetic polymorphisms. Finally, we applied inverse transformation of ranks to normality to obtain standard Normal distribution using *rntransform* function from *GenABEL* package for R[46].

After pre-processing we assessed the dependency between the glycan traits and such confounders as age, sex, and body-mass index (BMI). For age piece-wise relationships with glycan levels were found, which made it unjustified adding age in linear regression models as a confounder. Therefore, before further analysis we corrected glycan levels for age (through residuals) by segmented regression using 40–45 years as initial break-down points as implemented in *segmented* package for R[47]. The choice of the break-down points was done based on the observation of the correlation clouds for age and glycans followed by a bootstrap based search for "true" breakpoints. Depending on specific glycans, both BMI and sex exhibited remarkable (and significant) to negligible (and insignificant) linear relationships with glycan levels.

Linear mixed-models analysis. Because of the twin structure of the dataset, association analyses between disease status and glycan traits were performed using linear mixed models with *lme4* package for R with BMI and sex included as fixed covariates and variation in IgG glycan quantities between twin pairs as random effect. Also, for 36 participants a diagnosis of a major systemic inflammatory disorder was established, including rheumatoid arthritis, systemic lupus erythematosus, ulcerative colitis, and Crohn's disease. As people with such diagnoses normally undergo therapy with painkillers and anti-inflammatory medicine, and also the diseases have previously been found associated with variation in glycan levels, we included the diseases status as a fixed effect covariate in the analysis. The association was analysed for each glycan separately.

Weighted glycan "expression" networks. We used *WGCNA* package for R[48,49] to carry out an exploratory analysis of "network" dependencies between the glycan traits. The algorithm of the analysis is based on the estimation of correlations between the glycan levels across the dataset followed by extraction of relatively independent modules of correlated glycans. Glycan levels were adjusted for age, sex, BMI, and inflammatory disease status before the analysis. Signed networks algorithm was used which takes into account the direction of the correlation between glycans. The modules (represented by their eigenvalue estimated as first principal component for the glycans in every module) then were correlated with the pain phenotypes, including LBP and MRI trait LSUM. To estimate correlations between glycan modules and pain phenotypes we used point-biserial correlation coefficients and Pearson's correlation coefficients for qualitative and quantitative traits, respectively.

Discordant twins analysis. We compared the glycan levels in MZ and DZ twins discordant for LBP using paired t-test. Prior to the test, glycan levels were adjusted for age, sex, BMI, and inflammatory disease status.

The significance level consideration. There is an essential correlation between the glycan traits, many of which were derived from the original set of directly measured glycans. This complicates straightforward application of correction for multiple testing due to the violation of the requirement for the independence of the tests. Taking this into account, we estimated the effective number of independent statistical tests as of 19[50], which after Sidak's correction for multiple testing provided the significance level of 0.0027.

References

1. Brooks, P. M. The burden of musculoskeletal disease–a global perspective. *Clin Rheumatol* **25,** 778–781, doi: 10.1007/s10067-006-0240-3 (2006).
2. Andersson, G. B. Epidemiological features of chronic low-back pain. *Lancet* **354,** 581–585, doi: 10.1016/S0140-6736(99)01312-4 (1999).
3. Louw, Q. A., Morris, L. D. & Grimmer-Somers, K. The prevalence of low back pain in Africa: a systematic review. *BMC Musculoskelet Disord* **8,** 105, doi: 10.1186/1471-2474-8-105 (2007).
4. Steffens, D. *et al.* Does magnetic resonance imaging predict future low back pain? A systematic review. *Eur J Pain* **18,** 755–765, doi: 10.1002/j.1532-2149.2013.00427.x (2014).
5. Williams, F. M. & Sambrook, P. N. Neck and back pain and intervertebral disc degeneration: role of occupational factors. *Best Pract Res Clin Rheumatol* **25,** 69–79, doi: 10.1016/j.berh.2011.01.007 (2011).
6. Patel, A. A., Spiker, W. R., Daubs, M., Brodke, D. & Cannon-Albright, L. A. Evidence for an inherited predisposition to lumbar disc disease. *J Bone Joint Surg Am* **93,** 225–229, doi: 10.2106/JBJS.J.00276 (2011).
7. Livshits, G. *et al.* Lumbar disc degeneration and genetic factors are the main risk factors for low back pain in women: the UK Twin Spine Study. *Ann Rheum Dis* **70,** 1740–1745, doi: 10.1136/ard.2010.137836 (2011).
8. Kalichman, L. & Hunter, D. J. The genetics of intervertebral disc degeneration. Associated genes. *Joint Bone Spine* **75,** 388–396, doi: 10.1016/j.jbspin.2007.11.002 (2008).
9. Eskola, P. J. *et al.* Genetic association studies in lumbar disc degeneration: a systematic review. *PLos one* **7,** e49995, doi: 10.1371/journal.pone.0049995 (2012).
10. Song, Y. Q. *et al.* Lumbar disc degeneration is linked to a carbohydrate sulfotransferase 3 variant. *J Clin Invest* **123,** 4909–4917, doi: 10.1172/JCI69277 (2013).
11. Williams, F. M. *et al.* Novel genetic variants associated with lumbar disc degeneration in northern Europeans: a meta-analysis of 4600 subjects. *Ann Rheum Dis* **72,** 1141–1148, doi: 10.1136/annrheumdis-2012-201551 (2013).
12. Tajerian, M. *et al.* DNA methylation of SPARC and chronic low back pain. *Mol Pain* **7,** 65, doi: 10.1186/1744-8069-7-65 (2011).
13. Ohtsubo, K. & Marth, J. D. Glycosylation in cellular mechanisms of health and disease. *Cell* **126,** 855–867, doi: 10.1016/j.cell.2006.08.019 (2006).
14. Freeze, H. H. Genetic defects in the human glycome. *Nat Rev Genet* **7,** 537–551, doi: 10.1038/nrg1894 (2006).
15. Abbott, K. L. *et al.* Focused glycomic analysis of the N-linked glycan biosynthetic pathway in ovarian cancer. *Proteomics* **8,** 3210–3220, doi: 10.1002/pmic.200800157 (2008).
16. Nairn, A. V. *et al.* Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J Biol Chem* **283,** 17298–17313, doi: 10.1074/jbc.M801964200 (2008).
17. Gornik, O., Pavic, T. & Lauc, G. Alternative glycosylation modulates function of IgG and other proteins - implications on evolution and disease. *Biochim Biophys Acta* **1820,** 1318–1326, doi: 10.1016/j.bbagen.2011.12.004 (2012).
18. Malhotra, R. *et al.* Glycosylation changes of IgG associated with rheumatoid arthritis can activate complement via the mannose-binding protein. *Nat Med* **1,** 237–243 (1995).
19. Mihai, S. & Nimmerjahn, F. The role of Fc receptors and complement in autoimmunity. *Autoimmun Rev* **12,** 657–660, doi: 10.1016/j.autrev.2012.10.008 (2013).
20. Karsten, C. M. *et al.* Anti-inflammatory activity of IgG1 mediated by Fc galactosylation and association of FcgammaRIIB and dectin-1. *Nat Med* **18,** 1401–1406, doi: 10.1038/nm.2862 (2012).
21. Masuda, K. *et al.* Enhanced binding affinity for FcgammaRIIIa of fucose-negative antibody is sufficient to induce maximal antibody-dependent cellular cytotoxicity. *Mol Immunol* **44,** 3122–3131, doi: 10.1016/j.molimm.2007.02.005 (2007).
22. Umana, P., Jean-Mairet, J., Moudry, R., Amstutz, H. & Bailey, J. E. Engineered glycoforms of an antineuroblastoma IgG1 with optimized antibody-dependent cellular cytotoxic activity. *Nat Biotechnol* **17,** 176–180, doi: 10.1038/6179 (1999).
23. Horvath, S. *et al.* Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc Natl Acad Sci USA* **103,** 17402–17407, doi: 10.1073/pnas.0608396103 (2006).
24. Oldham, M. C., Horvath, S. & Geschwind, D. H. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc Natl Acad Sci USA* **103,** 17973–17978, doi: 10.1073/pnas.0605938103 (2006).

25. Fuller, T. F. *et al.* Weighted gene coexpression network analysis strategies applied to mouse weight. *Mamm Genome* **18,** 463–472, doi: 10.1007/s00335-007-9043-3 (2007).
26. Presson, A. P. *et al.* Integrated weighted gene co-expression network analysis with an application to chronic fatigue syndrome. *BMC Syst Biol* **2,** 95, doi: 10.1186/1752-0509-2-95 (2008).
27. Saris, C. G. *et al.* Weighted gene co-expression network analysis of the peripheral blood from Amyotrophic Lateral Sclerosis patients. *BMC Genomics* **10,** 405, doi: 10.1186/1471-2164-10-405 (2009).
28. van Eijk, K. R. *et al.* Genetic analysis of DNA methylation and gene expression levels in whole blood of healthy human subjects. *BMC Genomics* **13,** 636, doi: 10.1186/1471-2164-13-636 (2012).
29. Pucic, M. *et al.* 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, doi: 10.1074/mcp.M111.010090 (2011).
30. Lauc, G. *et al.* Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. *PLoS Genet* **9,** e1003225, doi: 10.1371/journal.pgen.1003225 (2013).
31. Vuckovic, F. *et al.* Systemic lupus erythematosus associates with the decreased immunosuppressive potential of the IgG glycome. *Arthritis Rheumatol*, doi: 10.1002/art.39273 (2015).
32. Gornik, O. & Lauc, G. Glycosylation of serum proteins in inflammatory diseases. *Dis Markers* **25,** 267–278 (2008).
33. Burke, J. G. *et al.* Intervertebral discs which cause low back pain secrete high levels of proinflammatory mediators. *J Bone Joint Surg Br* **84,** 196–201 (2002).
34. Sheth, T., Pitchumoni, C. S. & Das, K. M. Musculoskeletal manifestations in inflammatory bowel disease: a revisit in search of immunopathophysiological mechanisms. *J Clin Gastroenterol* **48,** 308–317, doi: 10.1097/MCG.0000000000000067 (2014).
35. Novokmet, M. *et al.* Changes in IgG and total plasma protein glycomes in acute systemic inflammation. *Sci Rep* **4,** 4347, doi: 10.1038/srep04347 (2014).
36. Moayyeri, A., Hammond, C. J., Hart, D. J. & Spector, T. D. The UK Adult Twin Registry (TwinsUK Resource). *Twin Res Hum Genet* **16,** 144–149, doi: 10.1017/thg.2012.89 (2013).
37. Smedley, J., Inskip, H., Cooper, C. & Coggon, D. Natural history of low back pain. A longitudinal study in nurses. *Spine (Phila Pa 1976)* **23,** 2422–2426 (1998).
38. MacGregor, A. J., Andrew, T., Sambrook, P. N. & Spector, T. D. Structural, psychological, and genetic influences on low back and neck pain: a study of adult female twins. *Arthritis Rheum* **51,** 160–167, doi: 10.1002/art.20236 (2004).
39. Livshits, G. *et al.* Evidence that bone mineral density plays a role in degenerative disc disease: the UK Twin Spine study. *Ann Rheum Dis* **69,** 2102–2106, doi: 10.1136/ard.2010.131441 (2010).
40. White, K. P., Speechley, M., Harth, M. & Ostbye, T. The London Fibromyalgia Epidemiology Study: the prevalence of fibromyalgia syndrome in London, Ontario. *J Rheumatol* **26,** 1570–1576 (1999).
41. Livshits, G. *et al.* An omics investigation into chronic widespread musculoskeletal pain reveals epiandrosterone sulfate as a potential biomarker. *Pain* **156,** 1845–1851, doi: 10.1097/j.pain.0000000000000200 (2015).
42. Paajanen, H., Lehto, I., Alanen, A., Erkintalo, M. & Komu, M. Diurnal fluid changes of lumbar discs measured indirectly by magnetic resonance imaging. *J Orthop Res* **12,** 509–514, doi: 10.1002/jor.1100120407 (1994).
43. Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal Chem* **78,** 4281–4290, doi: 10.1021/ac051632c (2006).
44. Chen, C. *et al.* Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods. *PLos one* **6,** e17238, doi: 10.1371/journal.pone.0017238 (2011).
45. Huffman, J. E. *et al.* Comparative performance of four methods for high-throughput glycosylation analysis of immunoglobulin G in genetic and epidemiological research. *Mol Cell Proteomics* **13,** 1598–1610, doi: 10.1074/mcp.M113.037465 (2014).
46. Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23,** 1294–1296, doi: 10.1093/bioinformatics/btm108 (2007).
47. Muggeo, V. M. Estimating regression models with unknown break-points. *Stat Med* **22,** 3055–3071, doi: 10.1002/sim.1545 (2003).
48. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9,** 559, doi: 10.1186/1471-2105-9-559 (2008).
49. Langfelder, P. & Horvath, S. Fast R Functions for Robust Correlations and Hierarchical Clustering. *J Stat Softw* **46,** i11, doi: 10.18637/jss.v046.i11 (2012).
50. Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* **95,** 221–227, doi: 10.1038/sj.hdy.6800717 (2005).

Acknowledgements

This work was carried out in framework of the European Community's Seventh Framework Programme funded PainOmics project (contract # 602736) and also supported by the European Community's Seventh Framework Programme MIMOmics (contract #305280), HTP-GlycoMet (contract #324400) and IntegraLife (contract #315997) grants and Arthritis Research UK (grant #7448). TwinsUK: the study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007–2013). The study also receives support from the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London.

Author Contributions

M.B.F. carried out statistical analysis; T.K., I.G., J.S., T.P. and M.S. performed glycan analysis; M.B.F., T.K., D.V., S.M.F., M.A., G.L. and F.M.K.W. were involved in interpretation of the results, drafting and revision of the manuscript; F.M.K.W. and G.L. conceived and designed the study. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: GL is founder and CEO of Genos – a private research organization that specialises in highthroughput glycomic analysis and has several patents in this field.

How to cite this article: Freidin, M. B. *et al.* The Association Between Low Back Pain and Composition of IgG Glycome. *Sci. Rep.* **6**, 26815; doi: 10.1038/srep26815 (2016).

5. GENERAL DISCUSSION

In contrast to proteins and nucleic acids, *N*-glycans are non-linear branched molecules. Because of their structural complexity and methodological difficulties associated with their analysis, the knowledge about the function of glycans and their role in disease mechanisms is lagging significantly behind the knowledge about the role of genes and proteins. However, as high-throughput analytical techniques for glycan analysis have been developed and more information about protein glycosylation emerges, it is becoming clear that glycosylation is strictly regulated and that glycan attachment to proteins is of great physiological significance [18]. Several large *N*-glycome population studies have been published until now [20–22,110] and they all revealed a high variability in the glycome composition between individuals. However, besides age, which significantly affected galactosylation, all other environmental factors individually accounted only for a small fraction of the observed variance [20], thus the main source of glycome variation between individuals is still unknown.

Using a classical study design with twins enrolled in the TwinsUK registry we have evaluated the heritable and non-heritable component of circulating IgG *N*-glycome and total plasma *N*-glycome. Variation in levels of 51 of the 76 IgG glycan traits studied was at least 50% heritable and only a small proportion of *N*-glycan traits had a low genetic contribution. Heritability of plasma *N*-glycome was also high, with half of the plasma glycan traits being at least 50% heritable (Table 1 in Supplemental data 2). Although glycans are shaped in a complex biosynthetic pathway [111] and were believed to be significantly affected by many environmental factors [112], we found a high contribution of the genetic component to *N*-glycome composition. It seems that genetic regulation of the IgG glycome (average heritability of all directly measured traits equals 53%) is slightly stronger than the regulation of the total plasma glycome (average heritability equals 47%). This makes sense regarding how IgG glycan structures can have a profound influence on IgG effector functions [48] and how proper IgG glycosylation is important for the function of the immune system [113]. Furthermore, plasma glycome is not affected only by the glycan amounts, but also by the concentration of each glycoprotein.

Even though most of the variation in IgG *N*-glycome in twins is contributed to genetic factors, interaction with the environment is also an important source of variability. Because some of this variation may be caused by epigenetic changes [114], we performed epigenome-wide association (EWA) analysis. It showed that methylation levels at some genes are also implicated in glycome composition, both in those with high heritabilities and those with a

lower genetic contribution. These genes have not been previously reported to have a role in IgG glycosylation. A similar observation was recently reported in a genome-wide association (GWA) study of the IgG glycome which identified 12 genes not previously known to be involved in IgG glycosylation [31], providing more evidence that IgG glycosylation is a very complex and tightly regulated process. Further cohorts with IgG *N*-glycome information, epigenetic and genetic data will be needed to explain the exact meaning and molecular pathways underlying this EWA and GWA hits.

By analysing IgG glycosylation in our population of twins, we also conducted the first study to investigate the potential role of IgG glycosylation in kidney function and identified 14 IgG glycan traits with high statistical significance associated with estimated glomerular filtration rate (eGFR). This was also validated in an independent subset of MZ twins discordant for renal disease, because MZ twins enable to carry out the ideal case–control study, as they are perfectly matched for age, genotype, family background, etc.

One of the prominently changed features of IgG glycosylation in chronic kidney disease (CKD) was the extent of galactosylation. We observed a higher risk of CKD in subjects with higher agalactosylated glycans and lower risk in those with higher galactosylated IgG. Decreased IgG galactosylation can lead to a more proinflammatory antibody response [48]. This change in IgG galactosylation has been reported in a number of inflammatory diseases [19], but also occurs with aging in the general population [65]. The decrease in galactosylation is not disease-specific, but is instead a general phenomenon associated with decreased immunosuppressive potential of circulating IgG.

Opposite to changes in galactosylation, the significant changes in sialylation have not been associated with many other diseases. Sialylation promotes recognition of IgG by DC-SIGN, which leads to increased expression of inhibitory FcγRIIB with consequent anti-inflammatory actions [50]. Sialylated IgG glycans are also believed to be the active fraction that harbors the anti-inflammatory potential of intravenous immunoglobulins (IVIg) therapy, which is used to suppress inflammation in a number of diseases [115]. Decreased IgG sialylation significantly reduces antiinflammatory activity of circulating IgG. In our population of twins, the sialylation was decreased in patients with CKD and sialylated glycans displayed a protective effect against CKD.

Approximately 18% of IgG glycans contain a bisecting GlcNAc, which significantly changes the structural properties of the glycan [22]. The effects of bisecting GlcNAc on

functional properties of IgG are not well understood [48], although increase in bisecting GlcNAc was reported to promote cell-mediated ADCC [59]. The addition of bisecting GlcNAc also inhibits α1,6-fucosyltransferase (FUT8) and the addition of core fucose to the glycan [60]. Fucosylation is crucial in many biological processes and inflammation in particular. IgG containing glycans that lack core fucose have over 50-fold increased affinity for FcγRIIIa and FcγRIIIb and are therefore much more efficient in activating ADCC than fucosylated glycoforms of the same molecule. It appears that the lack of core fucose and not the presence of bisecting GlcNAc has the most critical role in enhanced ADCC [61]. In our population of twins, the presence of bisecting GlcNAc on IgG glycans was associated with a higher risk of CKD. However, it is not clear how the modulation of ADCC could affect the renal damage in the onset of a nonautoimmune CKD.

To determine whether the changes in glycosylation in CKD were restricted to IgG or to a more general change in glycosylation of multiple proteins, we searched for an association between total plasma *N*-glycome and eGFR in a subset of 426 individuals. We found no difference in plasma glycosylation, suggesting that the effects we see here are likely direct effects of IgG glycosylation. However, the lack of association might also be due to power issues and so further study on larger sample size is needed to test this.

Using the WGCNA (weighted correlation network analysis) methodology [116], we carried out a network analysis for IgG glycan levels in twins to establish clusters of correlated glycans. WGCNA is a powerful methodology for revealing clusters (modules) of multiple omic traits and placing them into a biological context through the analysis of associations between the clusters and diseases or traits of interest. Using signed networks, we identified seven modules of correlated glycans. They reflected functionally related groups of glycans exhibiting overlapping biological activity. We also found associations between these clusters and pain phenotypes in twins with low back pain (LBP). We observed a positive correlation between pain phenotypes and "pro-ADCC" WGCNA glycan modules (high bisecting GlcNAc and low core fucose) and a negative correlation between pain phenotypes and "anti-ADCC" module (high core fucose, no bisecting GlcNAc). Although these correlations were weak, they may reflect some subtypes of LBP syndromes, which are related to inflammation [117] and in which ADCC contributes to the joint inflammation due to the decreased levels of core fucosylation and increased levels of bisecting GlcNAc in IgG glycans. This provides a new insight into understanding the complex pathophysiology of LBP and suggests glycan levels as a possible biomarker for inflammation-related subtypes of LBP.

Still, it is not clear whether changes in glycosylation are a consequence or a predisposition for the development of a disease. Despite the absence of a direct genetic template, the heritability of individual IgG glycans in twin population was very high (up to 80%), indicating that proinflammatory IgG glycome may also be a predisposition for the development of inflammatory diseases. This hypothesis is further supported by the fact that in rheumatoid arthritis the decrease in IgG galactosylation has been demonstrated to predate the onset of arthritis [118]. However, galactosylation of IgG is also dynamic and can change quite rapidly in acute inflammation [119]. Thus, it seems that both genetic and environmental factors strongly affect IgG galactosylation. This is also supported by the fact that glycome composition is associated with both genetic polymorphisms [29,31] and epigenetic modification on multiple genetic loci. It would be especially interesting to conduct longitudinal studies of the IgG glycome within an individual patient before and at different stages of the disease.

In our study on TwinsUK population we found association between IgG *N*-glycome and CKD and LBP. However, we did not find significant associations between total plasma proteins *N*-glycome and the diseases. Also, plasma glycan traits had a slightly lower heritability compared to IgG glycan traits. Interestingly, in previous studies, when glycome composition of isolated IgG was analysed in population, it varied more between individuals (more than three-fold) [22] compared to glycome composition of total plasma proteins [21]. Similarly, compared to GWAS study of the plasma glycome in 2705 individuals [29], the recent GWAS study of the IgG glycome in 2247 individuals [31] has identified five times more genetic loci with genome-wide significant associations. The variability of the composition of the plasma glycome derives from both variability in the composition of the plasma proteome and the variability in the glycosylation process. This difference implies that varying concentrations of plasma proteins and different structural and functional roles of the same glycans on different plasma proteins actually blur protein-specific regulation of glycosylation of individual proteins. By analysing total plasma glycome, glycans are probably averaged across the proteome which introduces considerable noise to the quantitation and interpretation of plasma glycan levels. Therefore, it may be beneficial to analyse glycosylation at a single protein level.

One of the limitations of this study was that, because of the novelty of the glycan phenotypes, we lack the replication for the epigenetic findings in an independent cohort. The fact that we find epigenome-wide significant hits on a relatively small sample suggests that epigenetic factors contribute to IgG glycan levels, although we cannot exclude false positive

results. Also, results from twin studies cannot be directly generalized to the general population because twins are not a random sample of the general population. Therefore, twin study results should be validated by demonstrating that twin samples are comparable to population-based samples of singletons. Furthermore, for historical reasons, the TwinsUK registry has considerably more female than male members.

The method that we used (HILIC-UPLC analysis) to analyse glycans is able to describe changes in galactosylation, sialylation, bisecting GlcNAc, and core fucosylation, but it cannot differentiate glycans released from Fc and Fab portions of IgG. Glycans from the Fc and Fab portions are known to be different, with Fab glycans having less core fucose and more galactose, sialic acid, and bisecting GlcNAc [46,98]. Nevertheless, only ~20% of the total IgG glycome originates from the Fab portion of IgG. In a small pilot of Fc-glycopeptides by nano-liquid chromatography tandem mass spectrometry [92] on 96 representative age-matched individuals from the extremes of the eGFR distribution, we found the same direction of effect with renal function for all but one. This suggests that the observed differences between patients and controls presumably originate from the Fc glycans.

This method for analyzing glycans is currently still not routinely used in clinics. However, it is foreseeable that in the near future, with advancing technology which allows high-throughput assessment of glycan profiles and with more information about protein glycosylation, glycan traits will become attractive and clinically feasible biomarkers. Moreover, increased knowledge about protein glycosylation could help to gain additional insights into the pathophysiology of disease and encourage development of new therapeutics.

6. CONCLUSIONS

In this thesis, heritable and non-heritable component of circulating IgG and total plasma protein *N*-glycome composition was evaluated, using a classical study design with twins enrolled in the TwinsUK registry. A high contribution of the genetic component to *N*-glycome composition was found. Variation in levels of 51 of the 76 IgG glycan traits studied was at least 50% heritable and only a small proportion of *N*-glycan traits had a low genetic contribution. Heritability of plasma *N*-glycome was also high, with half of the plasma glycan traits being at least 50% heritable. Average heritability of the IgG glycome was slightly larger than the average heritability of the total plasma glycome. Furthermore, epigenome-wide association (EWA) analysis showed that methylation levels at some genes are also implicated in glycome composition, both in those with high heritability and those with a lower genetic contribution.

By analysing IgG glycosylation in the population of twins, the first study to investigate the potential role of IgG glycosylation in kidney function was conducted. Fourteen glycan traits were associated with renal function in the discovery population ($P<6.5\times10^{-4}$) and remained significant after validation in an independent subset of MZ twins discordant for renal disease. Those glycan traits belong to three main glycosylation features: galactosylation, sialylation, and level of bisecting GlcNAc of the IgG glycans. These results indicate the role of IgG glycosylation in kidney function. They provide a novel insight into the pathophysiology of CKD and suggest glycan levels as a possible biomarker for CKD.

Using the WGCNA methodology, a network analysis for IgG glycan levels to establish clusters of correlated glycans was performed. Associations between these clusters and pain phenotypes in twins with low back pain were found. There was a weak positive correlation between pain phenotypes and "pro-ADCC" WGCNA glycan modules (high bisecting GlcNAc and low core fucose) and a weak negative correlation between pain phenotypes and "anti-ADCC" module (high core fucose, no bisecting GlcNAc). This suggests that some subtypes of LBP may be associated with an occult inflammation which is reflected by IgG glycan levels.

It appears that, even in the absence of a strict genetic template, variability in glycan levels is predominantly driven by genetic background and specific pathophysiological processes, which makes glycans to be promising candidates for biomarkers in many different diseases.

7. REFERENCE LIST

1. Kobata, A. A retrospective and prospective view of glycopathology. *Glycoconj. J.* **15,** 323–31 (1998).

2. Spiro, R. G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12,** 43R–56R (2002).

3. Dennis, J. W., Lau, K. S., Demetriou, M. & Nabi, I. R. Adaptive regulation at the cell surface by N-glycosylation. *Traffic* **10,** 1569–78 (2009).

4. Marth, J. D. & Grewal, P. K. Mammalian glycosylation in immunity. *Nat. Rev. Immunol.* **8,** 874–87 (2008).

5. Wells, L., Vosseller, K. & Hart, G. W. Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc. *Science* **291,** 2376–8 (2001).

6. Marek, K. W., Vijay, I. K. & Marth, J. D. A recessive deletion in the GlcNAc-1-phosphotransferase gene results in peri-implantation embryonic lethality. *Glycobiology* **9,** 1263–71 (1999).

7. Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* **1473,** 4–8 (1999).

8. Lebrilla, C. B. & An, H. J. The prospects of glycan biomarkers for the diagnosis of diseases. *Mol. Biosyst.* **5,** 17–20 (2009).

9. Lis, H. & Sharon, N. Protein glycosylation. Structural and functional aspects. *Eur. J. Biochem.* **218,** 1–27 (1993).

10. Varki, A. *et al.* Essentials of Glycobiology. (2009). at <http://www.ncbi.nlm.nih.gov/books/NBK1908/>

11. Council of Europe. in *Eur. Pharmacopoeia 7.0* 97–102 (Council of Europe, 2011).

12. Taniguchi, N. *et al. Handbook of Glycosyltransferases and Related Genes | Naoyuk Taniguchi | Springer*. (Springer Japan, 2014). at <http://www.springer.com/us/book/9784431542391>

13. Freeze, H. H. Genetic defects in the human glycome. *Nat. Rev. Genet.* **7,** 537–51 (2006).

14. Abbott, K. L. *et al.* Focused glycomic analysis of the N-linked glycan biosynthetic pathway in ovarian cancer. *Proteomics* **8,** 3210–20 (2008).

15. Nairn, A. V *et al.* Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J. Biol. Chem.* **283,** 17298–313 (2008).

16. Lauc, G., Vojta, A. & Zoldoš, V. Epigenetic regulation of glycosylation is the quantum mechanics of biology. *Biochim. Biophys. Acta* **1840,** 65–70 (2014).

17. Mariño, K., Saldova, R., Adamczyk, B. & Rudd, P. M. in *Carbohydr. Chem.* (ed. Pilar Rauter, A.) 57–93 (The Royal Society of Chemistry, 2012). doi:10.1039/9781849732765-00057

18. Moremen, K. W., Tiemeyer, M. & Nairn, A. V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.* **13,** 448–62 (2012).

19. Gornik, O. & Lauc, G. Glycosylation of serum proteins in inflammatory diseases. *Dis. Markers* **25,** 267–78 (2008).

20. Knežević, A. *et al.* Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma N-glycans. *Glycobiology* **20,** 959–69 (2010).

21. Knežević, A. *et al.* Variability, heritability and environmental determinants of human plasma N-glycome. *J. Proteome Res.* **8,** 694–701 (2009).

22. Pučić, M. *et al.* 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 (2011).

23. Ohtsubo, K. & Marth, J. D. Glycosylation in cellular mechanisms of health and disease. *Cell* **126,** 855–67 (2006).

24. Trbojević Akmačić, I. *et al.* Inflammatory bowel disease associates with proinflammatory potential of the immunoglobulin G glycome. *Inflamm. Bowel Dis.* **21,** 1237–47 (2015).

25. Axford, J. The impact of glycobiology on medicine. *Trends Immunol.* **22,** 237–9 (2001).

26. Miura, Y. *et al.* Change in N-Glycosylation of Plasma Proteins in Japanese

Semisupercentenarians. *PLoS One* **10,** e0142645 (2015).

27. Gornik, O. *et al.* Stability of N-glycan profiles in human plasma. *Glycobiology* **19,** 1547–53 (2009).

28. Dall'Olio, F. *et al.* N-glycomic biomarkers of biological aging and longevity: a link with inflammaging. *Ageing Res. Rev.* **12,** 685–98 (2013).

29. Lauc, G. *et al.* Genomics meets glycomics-the first GWAS study of human N-Glycome identifies HNF1α as a master regulator of plasma protein fucosylation. *PLoS Genet.* **6,** e1001256 (2010).

30. Huffman, J. E. *et al.* Polymorphisms in B3GAT1, SLC9A9 and MGAT5 are associated with variation within the human plasma N-glycome of 3533 European adults. *Hum. Mol. Genet.* **20,** 5000–11 (2011).

31. Lauc, G. *et al.* Loci Associated with N-Glycosylation of Human Immunoglobulin G Show Pleiotropy with Autoimmune Diseases and Haematological Cancers. *PLoS Genet.* **9,** e1003225 (2013).

32. Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era--concepts and misconceptions. *Nat. Rev. Genet.* **9,** 255–66 (2008).

33. Gielen, M. *et al.* Modeling genetic and environmental factors to increase heritability and ease the identification of candidate genes for birth weight: a twin study. *Behav. Genet.* **38,** 44–54 (2008).

34. Körner, C., Lehle, L. & von Figura, K. Carbohydrate-deficient glycoprotein syndrome type 1: correction of the glycosylation defect by deprivation of glucose or supplementation of mannose. *Glycoconj. J.* **15,** 499–505 (1998).

35. Ota, H. *et al.* Helicobacter pylori infection produces reversible glycosylation changes to gastric mucins. *Virchows Arch.* **433,** 419–26 (1998).

36. Tangvoranuntakul, P. *et al.* Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 12045–50 (2003).

37. Barišić, K., Lauc, G., Dumić, J., Pavlović, M. & Flögel, M. Changes of glycoprotein patterns in sera of humans under stress. *Eur. J. Clin. Chem. Clin. Biochem.* **34,** 97–101

(1996).

38. Clerc, F. *et al.* Human plasma protein N-glycosylation. *Glycoconj. J.* (2015).

39. Haltiwanger, R. S. & Lowe, J. B. Role of glycosylation in development. *Annu. Rev. Biochem.* **73,** 491–537 (2004).

40. Walt, D. *Transforming Glycoscience: A Roadmap for the Future*. (National Academies Press (US), 2012). at <http://www.ncbi.nlm.nih.gov/pubmed/23270009>

41. Arnold, J. N., Wormald, M. R., Sim, R. B., Rudd, P. M. & Dwek, R. A. The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annu. Rev. Immunol.* **25,** 21–50 (2007).

42. Pučić-Baković, M. Variability and heritability of immunoglobulin G glycosylation. (2013). at <https://bib.irb.hr/datoteka/670600.Doktorska_disertacija_Maja_Pucic_Bakovic.pdf>

43. Dashivets, T. *et al.* Multi-Angle Effector Function Analysis of Human Monoclonal IgG Glycovariants. *PLoS One* **10,** e0143520 (2015).

44. Huhn, C., Selman, M. H. J., Ruhaak, L. R., Deelder, A. M. & Wuhrer, M. IgG glycosylation analysis. *Proteomics* **9,** 882–913 (2009).

45. Jefferis, R. Glycosylation of recombinant antibody therapeutics. *Biotechnol. Prog.* **21,** 11–6 (2005).

46. van de Bovenkamp, F. S., Hafkenscheid, L., Rispens, T. & Rombouts, Y. The Emerging Importance of IgG Fab Glycosylation in Immunity. *J. Immunol.* **196,** 1435–1441 (2016).

47. Mimura, Y. *et al.* Role of oligosaccharide residues of IgG1-Fc in Fc gamma RIIb binding. *J. Biol. Chem.* **276,** 45539–47 (2001).

48. Gornik, O., Pavić, T. & Lauc, G. Alternative glycosylation modulates function of IgG and other proteins - implications on evolution and disease. *Biochim. Biophys. Acta* **1820,** 1318–26 (2012).

49. Malhotra, R. *et al.* Glycosylation changes of IgG associated with rheumatoid arthritis can activate complement via the mannose-binding protein. *Nat. Med.* **1,** 237–43

(1995).

50. Karsten, C. M. *et al.* Anti-inflammatory activity of IgG1 mediated by Fc galactosylation and association of FcγRIIB and dectin-1. *Nat. Med.* **18,** 1401–6 (2012).

51. Mihai, S. & Nimmerjahn, F. The role of Fc receptors and complement in autoimmunity. *Autoimmun. Rev.* **12,** 657–60 (2013).

52. Quast, I. *et al.* Sialylation of IgG Fc domain impairs complement-dependent cytotoxicity. *J. Clin. Invest.* **125,** 4160–70 (2015).

53. Anthony, R. M., Kobayashi, T., Wermeling, F. & Ravetch, J. V. Intravenous gammaglobulin suppresses inflammation through a novel T(H)2 pathway. *Nature* **475,** 110–3 (2011).

54. Boyd, P. N., Lines, A. C. & Patel, A. K. The effect of the removal of sialic acid, galactose and total carbohydrate on the functional activity of Campath-1H. *Mol. Immunol.* **32,** 1311–8 (1995).

55. Campbell, I. K. *et al.* Therapeutic effect of IVIG on inflammatory arthritis in mice is dependent on the Fc portion and independent of sialylation or basophils. *J. Immunol.* **192,** 5031–8 (2014).

56. Issekutz, A. C., Rowter, D., Miescher, S. & Käsermann, F. Intravenous IgG (IVIG) and subcutaneous IgG (SCIG) preparations have comparable inhibitory effect on T cell activation, which is not dependent on IgG sialylation, monocytes or B cells. *Clin. Immunol.* **160,** 123–32 (2015).

57. Masuda, K. *et al.* Enhanced binding affinity for FcgammaRIIIa of fucose-negative antibody is sufficient to induce maximal antibody-dependent cellular cytotoxicity. *Mol. Immunol.* **44,** 3122–31 (2007).

58. Davies, J. *et al.* 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,** 288–94 (2001).

59. Umaña, P., Jean-Mairet, J., Moudry, R., Amstutz, H. & Bailey, J. E. Engineered glycoforms of an antineuroblastoma IgG1 with optimized antibody-dependent cellular

cytotoxic activity. *Nat. Biotechnol.* **17,** 176–80 (1999).

60. Ferrara, C. *et al.* Modulation of therapeutic antibody effector functions by glycosylation engineering: influence of Golgi enzyme localization domain and co-expression of heterologous beta1, 4-N-acetylglucosaminyltransferase III and Golgi alpha-mannosidase II. *Biotechnol. Bioeng.* **93,** 851–61 (2006).

61. Shinkawa, T. *et al.* 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 (2003).

62. Mori, K. *et al.* Non-fucosylated therapeutic antibodies: the next generation of therapeutic antibodies. *Cytotechnology* **55,** 109–14 (2007).

63. Jung, S. T., Kang, T. H., Kelton, W. & Georgiou, G. Bypassing glycosylation: engineering aglycosylated full-length IgG antibodies for human therapy. *Curr. Opin. Biotechnol.* **22,** 858–67 (2011).

64. Zhang, A. *et al.* Understanding the conformational impact of chemical modifications on monoclonal antibodies with diverse sequence variation using hydrogen/deuterium exchange mass spectrometry and structural modeling. *Anal. Chem.* **86,** 3468–75 (2014).

65. Krištić, J. *et al.* Glycans are a novel biomarker of chronological and biological ages. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* **69,** 779–789 (2014).

66. Parekh, R., Roitt, I., Isenberg, D., Dwek, R. & Rademacher, T. Age-related galactosylation of the N-linked oligosaccharides of human serum IgG. *J. Exp. Med.* **167,** 1731–6 (1988).

67. Yamada, E., Tsukamoto, Y., Sasaki, R., Yagyu, K. & Takahashi, N. Structural changes of immunoglobulin G oligosaccharides with age in healthy human serum. *Glycoconj. J.* **14,** 401–5 (1997).

68. Parekh, R. B. *et al.* Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature* **316,** 452–7

69. Mehta, A. S. *et al.* Increased levels of galactose-deficient anti-Gal immunoglobulin G

in the sera of hepatitis C virus-infected individuals with fibrosis and cirrhosis. *J. Virol.* **82,** 1259–70 (2008).

70. Moore, J. S. *et al.* Increased levels of galactose-deficient IgG in sera of HIV-1-infected individuals. *AIDS* **19,** 381–9 (2005).

71. van de Geijn, F. E. *et al.* Immunoglobulin G galactosylation and sialylation are associated with pregnancy-induced improvement of rheumatoid arthritis and the postpartum flare: results from a large prospective cohort study. *Arthritis Res. Ther.* **11,** R193 (2009).

72. Parekh, R. B. *et al.* Galactosylation of IgG associated oligosaccharides: reduction in patients with adult and juvenile onset rheumatoid arthritis and relation to disease activity. *Lancet (London, England)* **1,** 966–9 (1988).

73. van Zeben, D. *et al.* Early agalactosylation of IgG is associated with a more progressive disease course in patients with rheumatoid arthritis: results of a follow-up study. *Br. J. Rheumatol.* **33,** 36–43 (1994).

74. Bond, A. *et al.* A detailed lectin analysis of IgG glycosylation, demonstrating disease specific changes in terminal galactose and N-acetylglucosamine. *J. Autoimmun.* **10,** 77–85 (1997).

75. Flögel, M., Lauc, G., Gornik, I. & Macek, B. Fucosylation and galactosylation of IgG heavy chains differ between acute and remission phases of juvenile chronic arthritis. *Clin. Chem. Lab. Med.* **36,** 99–102 (1998).

76. Dubé, R. *et al.* Agalactosyl IgG in inflammatory bowel disease: correlation with C-reactive protein. *Gut* **31,** 431–4 (1990).

77. Vučković, F. *et al.* Association of systemic lupus erythematosus with decreased immunosuppressive potential of the IgG glycome. *Arthritis Rheumatol. (Hoboken, N.J.)* **67,** 2978–89 (2015).

78. Perdivara, I., Peddada, S. D., Miller, F. W., Tomer, K. B. & Deterding, L. J. Mass spectrometric determination of IgG subclass-specific glycosylation profiles in siblings discordant for myositis syndromes. *J. Proteome Res.* **10,** 2969–78 (2011).

79. Kanoh, Y. *et al.* Changes in serum IgG oligosaccharide chains with prostate cancer

progression. *Anticancer Res.* **24,** 3135–9

80. Kanoh, Y. *et al.* Relationship between N-linked oligosaccharide chains of human serum immunoglobulin G and serum tumor markers with non-small cell lung cancer progression. *Anticancer Res.* **26,** 4293–7

81. Saldova, R. *et al.* Ovarian cancer is associated with changes in glycosylation in both acute-phase proteins and IgG. *Glycobiology* **17,** 1344–56 (2007).

82. Ruhaak, L. R. *et al.* The Serum Immunoglobulin G Glycosylation Signature of Gastric Cancer. *EuPA open proteomics* **6,** 1–9 (2015).

83. Kawaguchi-Sakita, N. *et al.* 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,** 1140–5 (2016).

84. Rakus, J. F. & Mahal, L. K. New technologies for glycomic analysis: toward a systematic understanding of the glycome. *Annu. Rev. Anal. Chem. (Palo Alto. Calif).* **4,** 367–92 (2011).

85. Cummings, R. D. The repertoire of glycan determinants in the human glycome. *Mol. Biosyst.* **5,** 1087–104 (2009).

86. Royle, L. *et al.* HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software. *Anal. Biochem.* **376,** 1–12 (2008).

87. Schwarzer, J., Rapp, E. & Reichl, U. N-glycan analysis by CGE-LIF: profiling influenza A virus hemagglutinin N-glycosylation during vaccine production. *Electrophoresis* **29,** 4203–14 (2008).

88. Ruhaak, L. R. *et al.* Optimized workflow for preparation of APTS-labeled N-glycans allowing high-throughput analysis of human plasma glycomes using 48-channel multiplexed CGE-LIF. *J. Proteome Res.* **9,** 6655–64 (2010).

89. Selman, M. H. J. *et al.* Immunoglobulin G glycopeptide profiling by matrix-assisted laser desorption ionization Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **82,** 1073–81 (2010).

90. Selman, M. H. J. *et al.* Fc specific IgG glycosylation profiling by robust nano-reverse phase HPLC-MS using a sheath-flow ESI sprayer interface. *J. Proteomics* **75,** 1318–29 (2012).

91. Reiding, K. R., Blank, D., Kuijper, D. M., Deelder, A. M. & Wuhrer, M. High-throughput profiling of protein N-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification. *Anal. Chem.* **86,** 5784–5793 (2014).

92. Huffman, J. E. *et al.* Comparative performance of four methods for high-throughput glycosylation analysis of immunoglobulin G in genetic and epidemiological research. *Mol. Cell. Proteomics* **13,** 1598–610 (2014).

93. Saldova, R. *et al.* Association of N-glycosylation with breast carcinoma and systemic features using high-resolution quantitative UPLC. *J. Proteome Res.* **13,** 2314–2327 (2014).

94. Holland, M. *et al.* 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,** 669–77 (2006).

95. Omtvedt, L. A. *et al.* Glycan analysis of monoclonal antibodies secreted in deposition disorders indicates that subsets of plasma cells differentially process IgG glycans. *Arthritis Rheum.* **54,** 3433–40 (2006).

96. Wuhrer, M. *et al.* Glycosylation profiling of immunoglobulin G (IgG) subclasses from human serum. *Proteomics* **7,** 4070–81 (2007).

97. Song, W. *et al.* N-glycoproteomics in plants: perspectives and challenges. *J. Proteomics* **74,** 1463–74 (2011).

98. Anumula, K. R. Quantitative determination of monosaccharides in glycoproteins by high-performance liquid chromatography with highly sensitive fluorescence detection. *Anal. Biochem.* **220,** 275–83 (1994).

99. Anumula, K. R. Advances in fluorescence derivatization methods for high-performance liquid chromatographic analysis of glycoprotein carbohydrates. *Anal. Biochem.* **350,** 1–23 (2006).

100. Ruhaak, L. R. *et al.* Glycan labeling strategies and their use in identification and

quantification. *Anal. Bioanal. Chem.* **397,** 3457–81 (2010).

101. Zaia, J. Mass spectrometry and the emerging field of glycomics. *Chem. Biol.* **15,** 881–92 (2008).

102. Guile, G. R., Rudd, P. M., Wing, D. R., Prime, S. B. & Dwek, R. A. A rapid high-resolution high-performance liquid chromatographic method for separating glycan mixtures and analyzing oligosaccharide profiles. *Anal. Biochem.* **240,** 210–26 (1996).

103. Bengtson, M.-B., Rønning, T., Vatn, M. H. & Harris, J. R. Irritable bowel syndrome in twins: genes and environment. *Gut* **55,** 1754–9 (2006).

104. Boomsma, D., Busjahn, A. & Peltonen, L. Classical twin studies and beyond. *Nat. Rev. Genet.* **3,** 872–82 (2002).

105. Ozaki, K., Toyoda, H., Iwama, N., Kubo, S. & Ando, J. Using non-normal SEM to resolve the ACDE model in the classical twin design. *Behav. Genet.* **41,** 329–39 (2011).

106. Tsankova, N., Renthal, W., Kumar, A. & Nestler, E. J. Epigenetic regulation in psychiatric disorders. *Nat. Rev. Neurosci.* **8,** 355–67 (2007).

107. Wong, A. H. C., Gottesman, I. I. & Petronis, A. Phenotypic differences in genetically identical organisms: the epigenetic perspective. *Hum. Mol. Genet.* **14 Spec No,** R11–8 (2005).

108. Fraga, M. F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 10604–9 (2005).

109. Vitaro, F., Brendgen, M. & Arseneault, L. The discordant MZ-twin method: One step closer to the holy grail of causality. *Int. J. Behav. Dev.* **33,** 376–382 (2009).

110. Ruhaak, L. R. *et al.* Decreased levels of bisecting GlcNAc glycoforms of IgG are associated with human longevity. *PLoS One* **5,** e12566 (2010).

111. Lauc, G., Rudan, I., Campbell, H. & Rudd, P. M. Complex genetic regulation of protein glycosylation. *Mol. Biosyst.* **6,** 329–35 (2010).

112. Lauc, G. & Zoldoš, V. Protein glycosylation--an evolutionary crossroad between genes and environment. *Mol. Biosyst.* **6,** 2373–9 (2010).

113. Nimmerjahn, F. & Ravetch, J. V. Fcgamma receptors as regulators of immune responses. *Nat. Rev. Immunol.* **8,** 34–47 (2008).

114. Zoldoš, V., Novokmet, M., Bečeheli, I. & Lauc, G. Genomics and epigenomics of the human glycome. *Glycoconj. J.* **30,** 41–50 (2013).

115. Schwab, I. & Nimmerjahn, F. Intravenous immunoglobulin therapy: how does IgG modulate the immune system? *Nat. Rev. Immunol.* **13,** 176–89 (2013).

116. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9,** 559 (2008).

117. Burke, J. G. *et al.* Intervertebral discs which cause low back pain secrete high levels of proinflammatory mediators. *J. Bone Joint Surg. Br.* **84,** 196–201 (2002).

118. Ercan, A. *et al.* Aberrant IgG galactosylation precedes disease onset, correlates with disease activity, and is prevalent in autoantibodies in rheumatoid arthritis. *Arthritis Rheum.* **62,** 2239–48 (2010).

119. Novokmet, M. *et al.* Changes in IgG and total plasma protein glycomes in acute systemic inflammation. *Sci Rep.* **4,** 4347 (2014).

120. Gornik, O., Keser, T. & Lauc, G. in *Sample Prep. Tech. Soil, Plant, Anim. Samples* (ed. Mičić, M.) (Humana Press, 2016).

121. Lauc, G., Pezer, M., Rudan, I. & Campbell, H. Mechanisms of disease: The human N-glycome. *Biochim. Biophys. Acta - Gen. Subj.* (2015).

122. Varki, A. *et al.* Symbol nomenclature for glycan representation. *Proteomics* **9,** 5398–9 (2009).

8. SUPPLEMENTAL DATA

8.1. **Supplemental data 1:** Glycosylation of Immunoglobulin G: Role of Genetic and Epigenetic Influences

Table S1. Description of the glycan codes.

Glycan Code	DESCRIPTION	FORMULA
GP1	*The percentage of FA1 glycan in total IgG glycans*	$GP1 / GP \times 100$
GP2	*The percentage of A2 glycan in total IgG glycans*	$GP2 / GP \times 100$
GP4	*The percentage of FA2 glycan in total IgG glycans*	$GP4 / GP \times 100$
GP5	*The percentage of M5 glycan in total IgG glycans*	$GP5 / GP \times 100$
GP6	*The percentage of FA2B glycan in total IgG glycans*	$GP6 / GP \times 100$
GP7	*The percentage of A2G1 glycan in total IgG glycans*	$GP7 / GP \times 100$
GP8	*The percentage of FA2[6]G1 glycan in total IgG glycans*	$GP8 / GP \times 100$
GP9	*The percentage of FA2[3]G1 glycan in total IgG glycans*	$GP9 / GP \times 100$
GP10	*The percentage of FA2[6]BG1 glycan in total IgG glycans*	$GP10 / GP \times 100$
GP11	*The percentage of FA2[3]BG1 glycan in total IgG glycans*	$GP11 / GP \times 100$
GP12	*The percentage of A2G2 glycan in total IgG glycans*	$GP12 / GP \times 100$
GP13	*The percentage of A2BG2 glycan in total IgG glycans*	$GP13 / GP \times 100$
GP14	*The percentage of FA2G2 glycan in total IgG glycans*	$GP14 / GP \times 100$
GP15	*The percentage of FA2BG2 glycan in total IgG glycans*	$GP15 / GP \times 100$
GP16	*The percentage of FA2G1S1 glycan in total IgG glycans*	$GP16 / GP \times 100$
GP17	*The percentage of A2G2S1 glycan in total IgG glycans*	$GP17/ GP \times 100$
GP18	*The percentage of FA2G2S1 glycan in total IgG glycans*	$GP18 / GP \times 100$
GP19	*The percentage of FA2BG2S1 glycan in total IgG glycans*	$GP19 / GP \times 100$
GP20	*Structure not determined*	$GP20 / GP \times 100$
GP21	*The percentage of A2G2S2 glycan in total IgG glycans*	$GP21 / GP \times 100$
GP22	*The percentage of A2BG2S2 glycan in total IgG glycans*	$GP22 / GP \times 100$
GP23	*The percentage of FA2G2S2 glycan in total IgG glycans*	$GP23 / GP \times 100$
GP24	*The percentage of FA2BG2S2 glycan in total IgG glycans*	$GP24 / GP \times 100$
FGS/(FG+FGS)	*The percentage of sialylation of fucosylated galactosylated structures without bisecting GlcNAc in total IgG glycans*	$\Sigma(GP16 + GP18 + GP23) / \Sigma(GP16 + GP18 + GP23 + GP8 + GP9 + GP14) \times 100$
FBGS/(FBG+FBGS)	*The percentage of sialylation of fucosylated galactosylated structures with bisecting GlcNAc in total IgG glycans*	$\Sigma(GP19 + GP24) / \Sigma(GP19 + GP24 + GP10 + GP11 +$

Glycan Code	DESCRIPTION	FORMULA
		GP15)× 100
FGS/(F+FG+FGS)	*The percentage of sialylation of all fucosylated structures without bisecting GlcNAc in total IgG glycans*	*∑(GP16 + GP18 + GP23) / ∑(GP16 + GP18 + GP23 + GP4 + GP8 + GP9 + GP14)× 100*
FBGS/(FB+FBG+FBGS)	*The percentage of sialylation of all fucosylated structures with bisecting GlcNAc in total IgG glycans*	*∑(GP19 + GP24) / ∑(GP19 + GP24 + GP6 + GP10 + GP11 + GP15)× 100*
FG1S1/(FG1+FG1S1)	*The percentage of monosialylation of fucosylated monogalactosylated structures in total IgG glycans*	*GP16 / ∑(GP16 + GP8 + GP9)× 100*
FG2S1/(FG2+FG2S1+FG2S2)	*The percentage of monosialylation of fucosylated digalactosylated structures in total IgG glycans*	*GP18 / ∑(GP18 + GP14 + GP23)× 100*
FG2S2/(FG2+FG2S1+FG2S2)	*The percentage of disialylation of fucosylated digalactosylated structures in total IgG glycans*	*GP23 / ∑(GP23 + GP14 + GP18)× 100*
FBG2S1/(FBG2+FBG2S1+FBG2S2)	*The percentage of monosialylation of fucosylated digalactosylated structures with bisecting GlcNAc in total IgG glycans*	*GP19 / ∑(GP19 + GP15 + GP24)× 100*
FBG2S2/(FBG2+FBG2S1+FBG2S2)	*The percentage of disialylation of fucosylated digalactosylated structures with bisecting GlcNAc in total IgG glycans*	*GP24 / ∑(GP24 + GP15 + GP19)× 100*
$F^{total}S1/F^{total}S2$	*Ratio of all fucosylated (+/- bisecting GlyNAc) monosialylated and disialylated structures in total IgG glycans*	*∑(GP16 + GP18 + GP19) / ∑(GP23 + GP24)*
FS1/FS2	*Ratio of fucosylated (without bisecting GlcNAc) monosialylated and disialylated structures in total IgG glycans*	*∑(GP16 + GP18) / GP23*
FBS1/FBS2	*Ratio of fucosylated (with bisecting GlcNAc) monosialylated and disialylated structures in total IgG glycans*	*GP19 / GP24*
FBS^{total}/FS^{total}	*Ratio of all fucosylated sialylated structures with and without bisecting GlcNAc*	*∑(GP19 + GP24) / ∑(GP16 + GP18 + GP23)*
FBS1/FS1	*Ratio of fucosylated monosialylated structures with and without bisecting GlcNAc*	*GP19 / ∑(GP16 + GP18)*
FBS1/(FS1+FBS1)	*The incidence of bisecting GlcNAc in all fucosylated monosialylated structures in total IgG glycans*	*GP19 / ∑(GP16 + GP18 + GP19)*
FBS2/FS2	*Ratio of fucosylated disialylated structures with and without bisecting GlcNAc*	*GP24 / GP23*
FBS2/(FS2+FBS2)	*The incidence of bisecting GlcNAc in all fucosylated disialylated structures in total IgG glycans*	*GP24 / ∑(GP23 + GP24)*
		GP = ∑(GP1:GP24)
$GP1^n$	*The percentage of FA1 glycan in total neutral IgG glycans (GP^n)*	*GP1 / GP^n× 100*
$GP2^n$	*The percentage of A2 glycan in total neutral IgG glycans (GP^n)*	*GP2 / GP^n× 100*
$GP4^n$	*The percentage of FA2 glycan in total neutral IgG glycans (GP^n)*	*GP4 / GP^n× 100*
$GP5^n$	*The percentage of M5 glycan in total neutral IgG glycans (GP^n)*	*GP5 / GP^n× 100*
$GP6^n$	*The percentage of FA2B glycan in total neutral IgG glycans (GP^n)*	*GP6 / GP^n× 100*

Glycan Code	DESCRIPTION	FORMULA
$GP7^n$	*The percentage of A2G1 glycan in total Ineutral IgG glycans (GP^n)*	$GP7 / GP^n \times 100$
$GP8^n$	*The percentage of FA2[6]G1 glycan in total neutral IgG glycans (GP^n)*	$GP8 / GP^n \times 100$
$GP9^n$	*The percentage of FA2[3]G1 glycan in total neutral IgG glycans (GP^n)*	$GP9 / GP^n \times 100$
$GP10^n$	*The percentage of FA2[6]BG1 glycan in total neutral IgG glycans (GP^n)*	$GP10 / GP^n \times 100$
$GP11^n$	*The percentage of FA2[3]BG1 glycan in total neutral IgG glycans (GP^n)*	$GP11 / GP^n \times 100$
$GP12^n$	*The percentage of A2G2 glycan in total neutral IgG glycans (GP^n)*	$GP12 / GP^n \times 100$
$GP13^n$	*The percentage of A2BG2 glycan in total neutral IgG glycans (GP^n)*	$GP13 / GP^n \times 100$
$GP14^n$	*The percentage of FA2G2 glycan in total neutral IgG glycans (GP^n)*	$GP14 / GP^n \times 100$
$GP15^n$	*The percentage of FA2BG2 glycan in total neutral IgG glycans (GP^n)*	$GP15 / GP^n \times 100$
$G0^n$	*The percentage of agalactosylated structures in total neutral IgG glycans*	$\sum(GP1^n : GP6^n)$
$G1^n$	*The percentage of monogalactosylated structures in total neutral IgG glycans*	$\sum(GP7^n : GP11^n)$
$G2^n$	*The percentage of digalactosylated structures in total neutral IgG glycans*	$\sum(GP12^n : GP15^n)$
$F^{n\ total}$	*The percentage of all fucosylated (+/- bisecting GlcNAc) structures in total neutral IgG glycans*	$\sum(GP1^n + GP4^n + GP5^n + GP6^n + GP8^n + GP9^n + GP10^n + GP11^n + GP14^n + GP15^n)$
$FG0^{n\ total}/G0^n$	*The percentage of fucosylation of agalactosylated structures*	$\sum(GP1^n + GP4^n + GP5^n + GP6^n) / G0^n \times 100$
$FG1^{n\ total}/G1^n$	*The percentage of fucosylation of monogalactosylated structures*	$\sum(GP8^n + GP9^n + GP10^n + GP11^n) / G1^n \times 100$
$FG2^{n\ total}/G2^n$	*The percentage of fucosylation of digalactosylated structures*	$\sum(GP14^n + GP15) / G2^n \times 100$
F^n	*The percentage of fucosylated (without bisecting GlcNAc) structures in total neutral IgG glycans*	$\sum(GP1^n + GP4^n + GP5^n + GP8^n + GP9^n + GP14^n)$
$FG0^n/G0^n$	*The percentage of fucosylation (without bisecting GlcNAc) of agalactosylated structures*	$\sum(GP1^n + GP4^n + GP5^n) / G0^n \times 100$
$FG1^n/G1^n$	*The percentage of fucosylation (without bisecting GlcNAc) of monogalactosylated structures*	$\sum(GP8^n + GP9^n) / G1^n \times 100$
$FG2^n/G2^n$	*The percentage of fucosylation (without bisecting GlcNAc) of digalactosylated structures*	$GP14^n / G2^n \times 100$
FB^n	*The percentage of fucosylated (with bisecting GlcNAc) structures in total neutral IgG glycans*	$\sum(GP6^n + GP10^n + GP11^n + GP15^n)$
$FBG0^n/G0^n$	*The percentage of fucosylation (with bisecting GlcNAc) of agalactosylated structures*	$GP6^n / G0^n \times 100$
$FBG1^n/G1^n$	*The percentage of fucosylation (with bisecting GlcNAc) of monogalactosylated structures*	$\sum(GP10^n + GP11^n) / G1^n \times 100$
$FBG2^n/G2^n$	*The percentage of fucosylation (with bisecting GlcNAc) of digalactosylated structures*	$GP15) / G2^n \times 100$
FB^n/F^n	*Ratio of fucosylated structures with and without bisecting GlcNAc*	$FB^n / F^n \times 100$

Glycan Code	DESCRIPTION	FORMULA
$FB^n/F^{n\ total}$	*The incidence of bisecting GlcNAc in all fucosylated structures in total neutral IgG glycans*	$FB^n/F^{n\ total} \times 100$
$F^n/(B^n + FB^n)$	*Ratio of fucosylated non-bisecting GlcNAc structures and all structures with bisecting GlcNAc*	$F^n/(GP13^n + FB^n)$
$B^n/(F^n + FB^n)$	*Ratio of structures with bisecting GlcNAc and all fucosylated structures (+/- bisecting GlcNAc)*	$GP13^n/(F^n + FB^n) \times 1000$
$FBG2^n/FG2^n$	*Ratio of fucosylated digalactosylated structures with and without bisecting GlcNAc*	$GP15^n/GP14^n$
$FBG2^n/(FG2^n + FBG2^n)$	*The incidence of bisecting GlcNAc in all fucosylated digalactosylated structures in total neutral IgG glycans*	$GP15^n/(GP14^n + GP15^n) \times 100$
$FG2^n/(BG2^n + FBG2^n)$	*Ratio of fucosylated digalactosylated non-bisecting GlcNAc structures and all digalactosylated structures with bisecting GlcNAc*	$GP14^n/(GP13^n + GP15^n)$
$BG2^n/(FG2^n + FBG2^n)$	*Ratio of digalactosylated structures with bisecting GlcNAc and all fucosylated digalactosylated structures (+/- bisecting GlcNAc)*	$GP15^n/(GP14^n + GP15^n) \times 1000$

$$GP^n = \Sigma(GP1^n:GP15^n)$$

Table S2. List of glycans associated with circulating levels of triglycerides (log) and of C-reactive protein (Bonferroni $P<7x10^{-4}$).

Phenotype	Glycan	Beta[95%CI]	P	h^2
tryglicerides	GP6	0.24[0.13,0.35]	$2.54 x10^{-5}$	0.75
tryglicerides	GP6n	0.24[0.13,0.35]	$3.30 x10^{-5}$	0.75
tryglicerides	GP18	-0.2[-0.32,-0.09]	$7.10 x10^{-4}$	0.73
tryglicerides	FG2n/(BG2n + FBG2n)	-0.21[-0.33,-0.09]	$7.16 x10^{-4}$	0.69
CRP	FG0n total/G0n	0.01[0,0.01]	$3.09 x10^{-4}$	0.52
CRP	GP8n	-0.01[-0.01,0]	$6.54 x10^{-4}$	0.80
CRP	GP7n	-0.01[-0.01,0]	$6.81x10^{-4}$	0.73

Table S3. List of loci associated with all glycans with high heritability (h^2>0.55).

Glycan trait	Gene 1	Gene 2	Gene 3	Gene 4
GP1	*SUV420H1*			
GP2	*IKZF1*	*ABCF2-SMARCD3*	*FUT8*	
GP4	*IL6ST-ANKRD55*			
GP5				
GP6	*ABCF2-SMARCD3*	*SYNGR1-TAB1-MGAT3-CACNA1I*		
GP7	*IKZF1*	*FUT8*		
GP8				
GP9				
GP10	*SMARCB1-DERL3*	*SYNGR1-TAB1-MGAT3-CACNA1I*		
GP11	*SMARCB1-DERL3*			
GP12	*FUT8*			
GP13				
GP17				
GP18	*ST6GAL1*	*B4GALT1*		
FGS/(F+FG+FGS)	*ST6GAL1*	*B4GALT1*		
FG2S2/(FG2+FG2S1+FG2S2)	*ST6GAL1*			
FBS2/FS2	*B4GALT1*	*SMARCB1-DERL3*	*SYNGR1-TAB1-MGAT3-CACNA1I*	
FBS2/(FS2+FBS2)	*B4GALT1*	*SMARCB1-DERL3*	*SYNGR1-TAB1-MGAT3-CACNA1I*	
GP1n	*SUV420H1*			
GP2n	*IKZF1*	*ABCF2-SMARCD3*	*FUT8*	
GP5n				
GP6n	*ABCF2-SMARCD3*	*SYNGR1-TAB1-MGAT3-CACNA1I*		
GP7n	*IKZF1*	*FUT8*		
GP8n				
GP9n				
GP10n	*SMARCB1-DERL3*	*SYNGR1-TAB1-MGAT3-CACNA1I*		
GP11n	*SMARCB1-DERL3*			
GP12n	*FUT8*			
GP13n				
G0n	*IL6ST-ANKRD55*			
G1n				
Fn total	*IKZF1*	*FUT8*		
FG0n total/G0n	*IKZF1*	*FUT8*		
FG1n total/G1n	*IKZF1*	*FUT8*		
FG2n total /G2n	*FUT8*			
Fn	*IKZF1*	*SMARCB1-DERL3*	*SYNGR1-TAB1-MGAT3-CACNA1I*	
FG0n/G0n	*IKZF1*	*FUT8*	*SMARCB1-DERL3*	*SYNGR1-TAB1-MGAT3-CACNA1I*
FG1n/G1n	*SMARCB1-DERL3*	*SYNGR1-TAB1-*		

		MGAT3-CACNA1I	
FG2n/G2n	*FUT8*		
FBn	*SMARCB1-DERL3*	*SYNGR1-TAB1-MGAT3-CACNA1I*	
FBG0n/G0n	*IKZF1*	*SMARCB1-DERL3*	*SYNGR1-TAB1-MGAT3-CACNA1I*
FBG1n/G1n	*SMARCB1-DERL3*	*SYNGR1-TAB1-MGAT3-CACNA1I*	
FBG2n/G2n	*SMARCB1-DERL3*		
FBn/Fn	*IKZF1*	*SMARCB1-DERL3*	*SYNGR1-TAB1-MGAT3-CACNA1I*
FBn/Fn total	*IKZF1*	*SMARCB1-DERL3*	*SYNGR1-TAB1-MGAT3-CACNA1I*
Fn/(Bn + FBn)	*IKZF1*	*SMARCB1-DERL3*	*SYNGR1-TAB1-MGAT3-CACNA1I*
Bn/(Fn + FBn) ‰			
FBG2n/FG2n	*SMARCB1-DERL3*		
FBG2n /(FG2n + FBG2n)	*SMARCB1-DERL3*		
FG2n/(BG2n + FBG2n)	*SMARCB1-DERL3*		
BG2n/(FG2n + FBG2n) ‰			

8.2. Supplemental data 2: Table 1

The amount of variance attributed to: additive genetic factors (A or heritability), common/shared environmental factors (C) and unique environmental factors (E) for total plasma *N*-glycome (data from 440 monozygotic and 610 dizygotic female twins were analysed, data not published).

Glycan peak[a]	Glycan structure[b]	Best model	A	C	E
GP1	FA2	ACE	0.53	0.24	0.23
GP2	M5 FA2B	ACE	0.54	0.24	0.21
GP3	A2[6]BG1	AE	0.49		0.51
GP4	FA2[6]G1	AE	0.73		0.27
GP5	FA2[3]G1	AE	0.77		0.23
GP6	FA2[6]BG1	AE	0.81		0.19
GP7	M6 FA2[3]BG1	AE	0.53		0.47
GP8	A2G2	ACE	0.55	0.14	0.30
GP9	A2BG2	AE	0.61		0.39
GP10+11	FA2G2	ACE	0.44	0.32	0.24
GP12	FA2BG2	ACE	0.11	0.29	0.59
GP13	A2[3]BG1S[3]1 A2[3]BG1S[6]1	ACE	0.46	0.21	0.33
GP14	FA2[3]G1S[3]1 FA2[3]G1S[6]1	AE	0.71		0.29
GP15	A2G2S[6]1 A2G2S[3]1	ACE	0.18	0.15	0.68
GP16	A2BG2S[6]1	AE	0.71		0.29
GP17	FA2G2S[6]1	ACE	0.49	0.16	0.36
GP18	FA2G2S[6]1	ACE	0.62	0.20	0.18
GP19	FA2BG2S[3]1 FA2BG2S[6]1	AE	0.72		0.28
GP20+21	A2G2S[3,6]2	AE	0.63		0.37
GP22	M9 A3G3S[3]1 A3G3S[6]1	ACE	0.25	0.31	0.44
GP23	A2G2S[3,6]2	AE	0.67		0.33
GP24	A2BG2S[3,6]2	AE	0.65		0.35
GP25	FA2G2S[3,6]2	AE	0.75		0.25
GP26	FA2BG2S[3,6]2 FA2BG2S[6,6]2	ACE	0.47	0.34	0.19
GP27+28	A3G3S[3,6]2 A3BG3S[3,6]2	AE	0.71		0.29
GP29	A3G3S[3,3]2	ACE	0.36	0.21	0.43
GP30	A3G3S[3,3,3]3	ACE	0.40	0.15	0.46
GP31+32	A3G3S[3,3,6]3 FA3G3S[3,3,3]3	AE	0.49		0.51
GP33	A3G3S[3,3,6]3	AE	0.53		0.47
GP34	FA3G3S[3,3,6]3 FA3G3S[3,6,6]3	AE	0.48		0.52
GP35	A3F1G3S[3,3,3]3	ACE	0.17	0.40	0.43
GP36	A4G4S[3,3,3]3	ACE	0.39	0.24	0.37

GP37	A4G4S[3,3,6]3 A4G4S[3,6,6]3	ACE	0.25	0.47	0.28
GP38	A4F1G3S[3,3,3]3 A4F1G3S[3,3,6]3 A4F1G3S[3,6,6]3	ACE	0.51	0.17	0.32
GP39	A4G4S[3,3,3,3]4	CE		0.48	0.52
GP40	A4G4S[3,3,3,6]4	ACE	0.15	0.32	0.53
GP41	A4G4S[3,3,3,6]4	CE		0.38	0.62
GP42	A4F1G4S[3,3,3,6]4	ACE	0.12	0.28	0.59

[a]42 glycan peaks obtained by UPLC analysis of released *N*-glycans from total plasma proteins. [b]The most abundant glycan structures in each glycan peak - structure abbreviations: all *N*-glycans have two core GlcNAcs; F at the start of the abbreviation indicates a core fucose α1-6 linked to the inner GlcNAc; M*x*, number (*x*) of mannose on core GlcNAcs; A*x*, number of antenna (GlcNAc) on trimannosyl core; A2, biantennary with both GlcNAcs as *β*1-2 linked; A3, triantennary with a GlcNAc linked *β*1-2 to both mannose and the third GlcNAc linked *β*1-4 to the α1-3 linked mannose; A4, GlcNAcs linked as A3 with additional GlcNAc *β*1-6 linked to α1-6 mannose; B, bisecting GlcNAc linked *β*1-4 to *β*1-3 mannose; G*x*, number (*x*) of α1-4 linked galactose on antenna; [3]G1 and [6]G1 indicates that the galactose is on the antenna of the α1-3 or α1-6 mannose; F(*x*), number (*x*) of fucose linked α1-3 to antenna GlcNAc; S*x*, number (*x*) of sialic acids linked to galactose; the numbers 3 or 6 or in parentheses after S indicate whether the sialic acid is in an α2-3 or α2-6 linkage.

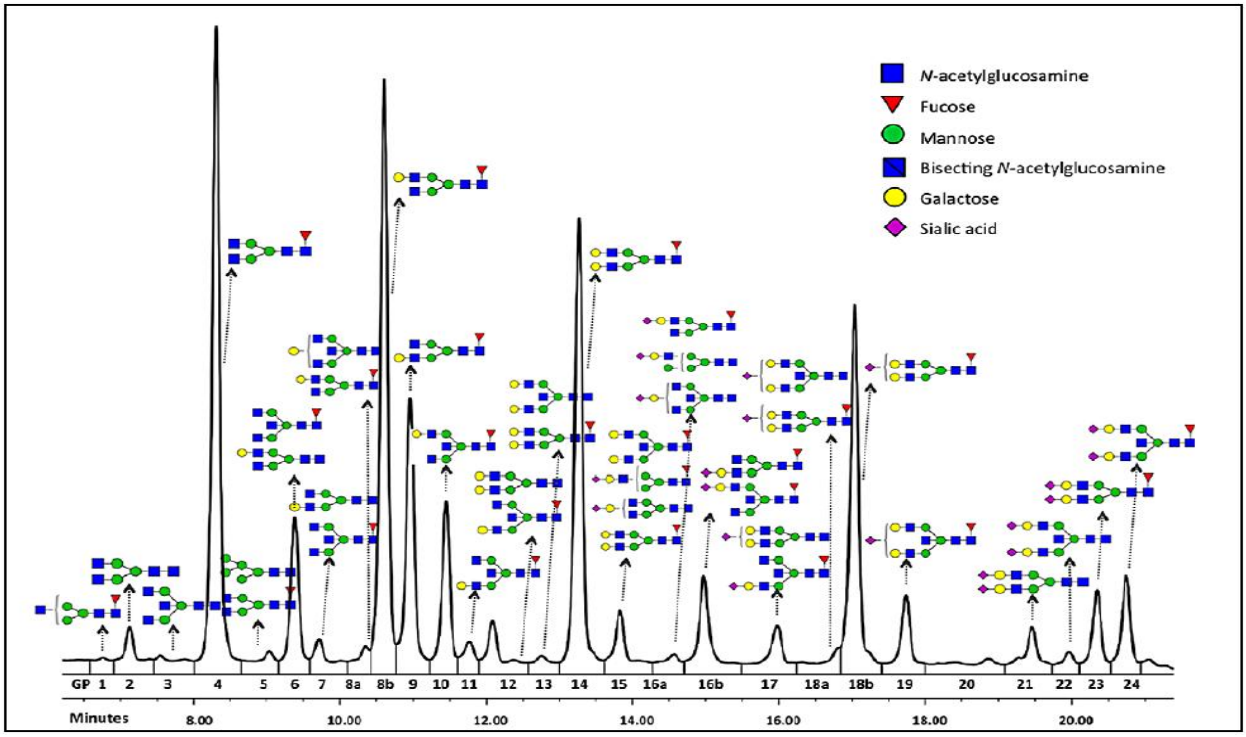
8.3. **Supplemental data 3:** Glycosylation Profile of IgG in Moderate Kidney Dysfunction

Supplementary Table 1. Glycan traits and their association with eGFR. Description of 24 quantitative IgG glycosylation traits and 52 derived traits and association between all tested glycans and derived traits with CKD status and eGFR.

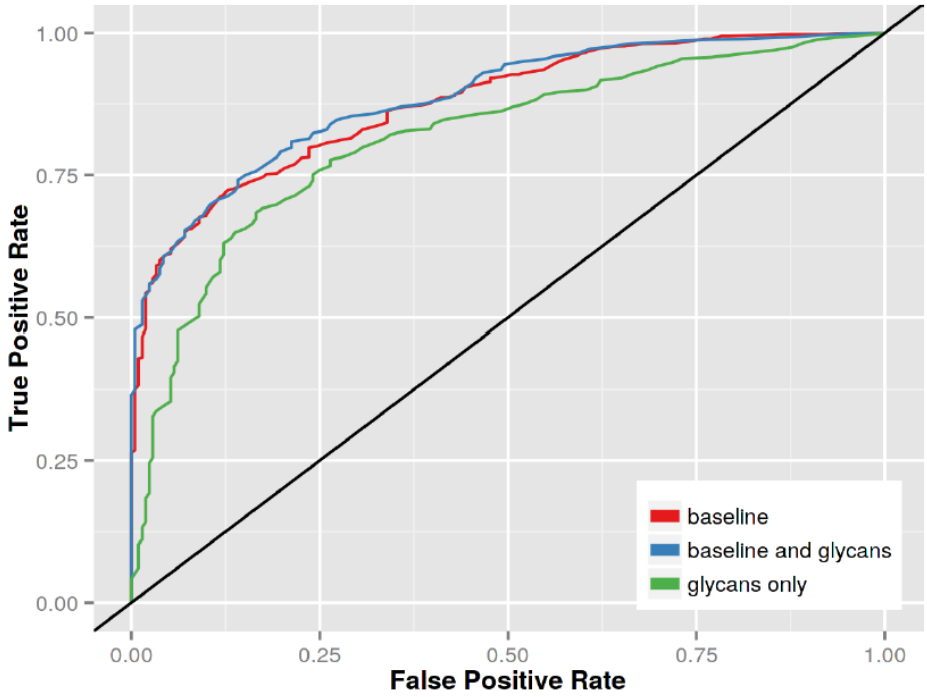
				eGFR	
GROUP	**Glycan**	**DESCRIPTION***	**FORMULA***	**Beta [95%CI]**	**P**
Total IgG glycans (neutral + charged)	GP1	*The percentage of FA1 glycan in total IgG glycans*	*GP1 / GP* 100*	-0.26 [-0.74:0.22]	$2.89x10^{-1}$
	GP2	*The percentage of A2 glycan in total IgG glycans*	*GP2 / GP* 100*	-0.90 [-1.42:-0.38]	$6.28x10^{-4}$
	GP4	*The percentage of FA2 glycan in total IgG glycans*	*GP4 / GP* 100*	-0.60 [-1.14:-0.06]	$2.87x10^{-2}$
	GP5	*The percentage of M5 glycan in total IgG glycans*	*GP5 / GP* 100*	-0.24 [-0.73:0.25]	$3.30x10^{-1}$
	GP6	*The percentage of FA2B glycan in total IgG glycans*	*GP6 / GP* 100*	-1.14 [-1.71:-0.57]	$8.90x10^{-5}$
	GP7	*The percentage of A2G1 glycan in total IgG glycans*	*GP7 / GP* 100*	-0.53 [-1.00:-0.06]	$2.78x10^{-2}$
	GP8	*The percentage of FA2[6]G1 glycan in total IgG glycans*	*GP8 / GP* 100*	0.46 [-0.02:0.93]	$5.89x10^{-2}$
	GP9	*The percentage of FA2[3]G1 glycan in total IgG glycans*	*GP9 / GP* 100*	0.07 [-0.39:0.53]	$7.69x10^{-1}$
	GP10	*The percentage of FA2[6]BG1 glycan in total IgG glycans*	*GP10 / GP* 100*	-0.13 [-0.60:0.34]	$5.86x10^{-1}$
	GP11	*The percentage of FA2[3]BG1 glycan in total IgG glycans*	*GP11 / GP* 100*	-0.38 [-0.88:0.11]	$1.30x10^{-1}$
	GP12	*The percentage of A2G2 glycan in total IgG glycans*	*GP12 / GP* 100*	0.18 [-0.30:0.66]	$4.58x10^{-1}$
	GP13	*The percentage of A2BG2 glycan in total IgG glycans*	*GP13 / GP* 100*	0.33 [-0.16:0.82]	$1.91x10^{-1}$
	GP14	*The percentage of FA2G2 glycan in total IgG glycans*	*GP14 / GP* 100*	1.46 [0.85:2.07]	$2.92x10^{-6}$
	GP15	*The percentage of FA2BG2 glycan in total IgG glycans*	*GP15 / GP* 100*	0.70 [0.17:1.23]	$9.47x10^{-3}$
	GP16	*The percentage of FA2G1S1 glycan in total IgG glycans*	*GP16 / GP * 100*	0.51 [0.03:0.99]	$3.66x10^{-2}$
	GP17	*The percentage of A2G2S1 glycan in total IgG glycans*	*GP17/ GP * 100*	-0.09 [-0.55:0.37]	$7.10x10^{-1}$
	GP18	*The percentage of FA2G2S1 glycan in total IgG glycans*	*GP18 / GP * 100*	1.48 [0.89:2.07]	$8.60x10^{-7}$
	GP19	*The percentage of FA2BG2S1 glycan in total IgG glycans*	*GP19 / GP * 100*	-0.11 [-0.57:0.35]	$6.37x10^{-1}$
	GP20+GP21		*(GP20+GP21)/GP * 100*	0.14 [-0.14:0.42]	$3.38x10^{-1}$
	GP22	*The percentage of A2BG2S2 glycan in total IgG glycans*	*GP22 / GP * 100*	0.13 [-0.32:0.58]	$5.67x10^{-1}$
	GP23	*The percentage of FA2G2S2 glycan in total IgG glycans*	*GP23 / GP * 100*	0.69 [0.19:1.18]	$6.63x10^{-3}$
	GP24	*The percentage of FA2BG2S2 glycan in total IgG glycans*	*GP24 / GP * 100*	0.08 [-0.38:0.53]	$7.36x10^{-1}$
Total IgG glycans - derived parameters	FGS/(FG+FGS)	*The percentage of sialylation of fucosylated galactosylated structures without bisecting GlcNAc in total IgG glycans*	*SUM(GP16 + GP18 + GP23) / SUM(GP16 + GP18 + GP23 + GP8 + GP9 + GP14)* 100*	0.76 [0.25:1.27]	$3.35x10^{-3}$
	FBGS/(FBG+FBGS)	*The percentage of sialylation of fucosylated galactosylated structures with bisecting GlcNAc in total IgG glycans*	*SUM(GP19 + GP24) / SUM(GP19 + GP24 + GP10 + GP11 + GP15)* 100*	-0.06 [-0.52:0.41]	$8.11x10^{-1}$
	FGS/(F+FG+FGS)	*The percentage of sialylation of all fucosylated structures without bisecting GlcNAc in total IgG glycans*	*SUM(GP16 + GP18 + GP23) / SUM(GP16 + GP18 + GP23 + GP4 + GP8 + GP9 + GP14)* 100*	1.01 [0.46:1.56]	$2.96x10^{-4}$
	FBGS/(FB+FBG+FBGS)	*The percentage of sialylation of all fucosylated structures with bisecting GlcNAc in total IgG glycans*	*SUM(GP19 + GP24) / SUM(GP19 + GP24 + GP6 + GP10 + GP11 + GP15)* 100*	0.24 [-0.23:0.72]	$3.16x10^{-1}$
	FG1S1/(FG1+FG1S1)	*The percentage of monosialylation of fucosylated monogalactosylated structures in total IgG glycans*	*GP16 / SUM(GP16 + GP8 + GP9)* 100*	0.20 [-0.28:0.67]	$4.19x10^{-1}$
	FG2S1/(FG2+FG2S1+FG2S2)	*The percentage of monosialylation of fucosylated digalactosylated structures in total IgG glycans*	*GP18 / SUM(GP18 + GP14 + GP23)* 100*	0.27 [-0.22:0.77]	$2.82x10^{-1}$
	FG2S2/(FG2+FG2S1+FG2S2)	*The percentage of disialylation of fucosylated digalactosylated structures in total IgG glycans*	*GP23 / SUM(GP23 + GP14 + GP18)* 100*	-0.27 [-0.75:0.21]	$2.72x10^{-1}$
	FBG2S1/(FBG2+FBG2S1+FBG2S2)	*The percentage of monosialylation of fucosylated digalactosylated structures with bisecting GlcNAc in total IgG glycans*	*GP19 / SUM(GP19 + GP15 + GP24)* 100*	-0.63 [-1.12:-0.13]	$1.25x10^{-2}$
	FBG2S2/(FBG2+FBG2S1+FBG2S2)	*The percentage of disialylation of fucosylated digalactosylated structures with bisecting GlcNAc in total IgG glycans*	*GP24 / SUM(GP24 + GP15 + GP19)* 100*	-0.11 [-0.59:0.37]	$6.59x10^{-1}$
	$F^{total}S1/F^{total}S2$	*Ratio of all fucosylated (+/- bisecting GlyNAc) monosialylated and disialylated structures in total IgG glycans*	*SUM(GP16 + GP18 + GP19) / SUM(GP23 + GP24)*	0.49 [0.00:0.98]	$4.77x10^{-2}$
	FS1/FS2	*Ratio of fucosylated (without bisecting GlcNAc) monosialylated and disialylated structures in total IgG*	*SUM(GP16 + GP18) / GP23*	0.17 [-0.30:0.64]	$4.71x10^{-1}$

		glycans			
	FBS1/FBS2	*Ratio of fucosylated (with bisecting GlcNAc) monosialylated and disialylated structures in total IgG glycans*	*GP19 / GP24*	-0.30 [-0.78:0.17]	$2.09x10^{-1}$
	FBS^{total}/FS^{total}	*Ratio of all fucosylated sialylated structures with and without bisecting GlcNAc*	*SUM(GP19 + GP24) / SUM(GP16 + GP18 + GP23)*	-1.07 [-1.60:-0.54]	$8.21x10^{-5}$
	FBS1/FS1	*Ratio of fucosylated monosialylated structures with and without bisecting GlcNAc*	*GP19 / SUM(GP16 + GP18)*	-1.12 [-1.65:-0.59]	$3.48x10^{-5}$
	FBS1/(FS1+FBS1)	*The incidence of bisecting GlcNAc in all fucosylated monosialylated structures in total IgG glycans*	*GP19 / SUM(GP16 + GP18 + GP19)*	-1.10 [-1.63:-0.57]	$4.63x10^{-5}$
	FBS2/FS2	*Ratio of fucosylated disialylated structures with and without bisecting GlcNAc*	*GP24 / GP23*	-0.90 [-1.42:-0.37]	$8.45x10^{-4}$
	FBS2/(FS2+FBS2)	*The incidence of bisecting GlcNAc in all fucosylated disialylated structures in total IgG glycans*	*GP24 / SUM(GP23 + GP24)*	-0.91 [-1.44:-0.37]	$8.47x10^{-4}$
Neutral IgG glycans	$GP1^n$	*The percentage of FA1 glycan in total neutral IgG glycans (GP^n)*	*$GP1 / GP^n * 100$*	-0.32 [-0.81:0.16]	$1.92x10^{-1}$
	$GP2^n$	*The percentage of A2 glycan in total neutral IgG glycans (GP^n)*	*$GP2 / GP^n * 100$*	-0.91 [-1.42:-0.40]	$5.02x10^{-4}$
	$GP4^n$	*The percentage of FA2 glycan in total neutral IgG glycans (GP^n)*	*$GP4 / GP^n * 100$*	-0.90 [-1.47:-0.33]	$2.04x10^{-3}$
	$GP5^n$	*The percentage of M5 glycan in total neutral IgG glycans (GP^n)*	*$GP5 / GP^n * 100$*	-0.26 [-0.75:0.22]	$2.87x10^{-1}$
	$GP6^n$	*The percentage of FA2B glycan in total neutral IgG glycans (GP^n)*	*$GP6 / GP^n * 100$*	-1.39 [-1.98:-0.80]	$3.56x10^{-6}$
	$GP7^n$	*The percentage of A2G1 glycan in total Ineutral IgG glycans (GP^n)*	*$GP7 / GP^n * 100$*	-0.52 [-0.99:-0.05]	$2.87x10^{-2}$
	$GP8^n$	*The percentage of FA2[6]G1 glycan in total neutral IgG glycans (GP^n)*	*$GP8 / GP^n * 100$*	0.53 [0.02:1.04]	$3.98x10^{-2}$
	$GP9^n$	*The percentage of FA2[3]G1 glycan in total neutral IgG glycans (GP^n)*	*$GP9 / GP^n * 100$*	-0.02 [-0.48:0.45]	$9.46x10^{-1}$
	$GP10^n$	*The percentage of FA2[6]BG1 glycan in total neutral IgG glycans (GP^n)*	*$GP10 / GP^n * 100$*	-0.21 [-0.68:0.26]	$3.81x10^{-1}$
	$GP11^n$	*The percentage of FA2[3]BG1 glycan in total neutral IgG glycans (GP^n)*	*$GP11 / GP^n * 100$*	-0.50 [-0.99:-0.01]	$4.71x10^{-2}$
	$GP12^n$	*The percentage of A2G2 glycan in total neutral IgG glycans (GP^n)*	*$GP12 / GP^n * 100$*	0.13 [-0.35:0.61]	$5.93x10^{-1}$
	$GP13^n$	*The percentage of A2BG2 glycan in total neutral IgG glycans (GP^n)*	*$GP13 / GP^n * 100$*	0.20 [-0.29:0.69]	$4.20x10^{-1}$
	$GP14^n$	*The percentage of FA2G2 glycan in total neutral IgG glycans (GP^n)*	*$GP14 / GP^n * 100$*	1.29 [0.68:1.90]	$3.06x10^{-5}$
	$GP15^n$	*The percentage of FA2BG2 glycan in total neutral IgG glycans (GP^n)*	*$GP15 / GP^n * 100$*	0.51 [-0.01:1.03]	$5.55x10^{-2}$
Neutral IgG glycans - derived parameters	$G0^n$	*The percentage of agalactosylated structures in total neutral IgG glycans*	*$SUM(GP1^n: GP6^n)$*	-1.16 [-1.76:-0.56]	$1.52x10^{-4}$
	$G1^n$	*The percentage of monogalactosylated structures in total neutral IgG glycans*	*$SUM(GP7^n: GP11^n)$*	0.36 [-0.12:0.84]	$1.45x10^{-1}$
	$G2^n$	*The percentage of digalactosylated structures in total neutral IgG glycans*	*$SUM(GP12^n: GP15^n)$*	1.20 [0.60:1.80]	$8.81x10^{-5}$
	$F^{n\ total}$	*The percentage of all fucosylated (+/- bisecting GlcNAc) structures in total neutral IgG glycans*	*$SUM(GP1^n + GP4^n + GP5^n + GP6^n + GP8^n + GP9^n + GP10^n + GP11^n + GP14^n + GP15^n)$*	0.31 [-0.16:0.77]	$1.93x10^{-1}$
	$FG0^{n\ total}/G0^n$	*The percentage of fucosylation of agalactosylated structures*	*$SUM(GP1^n + GP4^n + GP5^n + GP6^n) / G0^n * 100$*	0.49 [0.01:0.98]	$4.67x10^{-2}$
	$FG1^{n\ total}/G1^n$	*The percentage of fucosylation of monogalactosylated structures*	*$SUM(GP8^n + GP9^n + GP10^n + GP11^n) / G1^n * 100$*	0.53 [0.05:1.00]	$2.89x10^{-2}$
	$FG2^{n\ total}/G2^n$	*The percentage of fucosylation of digalactosylated structures*	*$SUM(GP14^n + GP15) / G2^n * 100$*	0.44 [-0.04:0.92]	$7.54x10^{-2}$
	F^n	*The percentage of fucosylated (without bisecting GlcNAc) structures in total neutral IgG glycans*	*$SUM(GP1^n + GP4^n + GP5^n + GP8^n + GP9^n + GP14^n)$*	0.58 [0.09:1.07]	$2.15x10^{-2}$
	$FG0^n/G0^n$	*The percentage of fucosylation (without bisecting GlcNAc) of agalactosylated structures*	*$SUM(GP1^n + GP4^n + GP5^n) / G0^n * 100$*	0.36 [-0.11:0.83]	$1.29x10^{-1}$

	$FG1^n/G1^n$	*The percentage of fucosylation (without bisecting GlcNAc) of monogalactosylated structures*	*$SUM(GP8^n + GP9^n) / G1^n * 100$*	0.36 [-0.12:0.85]	$1.39x10^{-1}$
	$FG2^n/G2^n$	*The percentage of fucosylation (without bisecting GlcNAc) of digalactosylated structures*	*$GP14^n / G2^n * 100$*	0.73 [0.23:1.24]	$4.44x10^{-3}$
	FB^n	*The percentage of fucosylated (with bisecting GlcNAc) structures in total neutral IgG glycans*	*$SUM(GP6^n + GP10^n + GP11^n + GP15^n)$*	-0.68 [-1.19:-0.17]	$8.51x10^{-3}$
	$FBG0^n/G0^n$	*The percentage of fucosylation (with bisecting GlcNAc) of agalactosylated structures*	*$GP6^n / G0^n * 100$*	-0.42 [-0.90:0.05]	$7.88x10^{-2}$
	$FBG1^n/G1^n$	*The percentage of fucosylation (with bisecting GlcNAc) of monogalactosylated structures*	*$SUM(GP10^n + GP11^n) / G1^n * 100$*	-0.36 [-0.85:0.12]	$1.40x10^{-1}$
	$FBG2^n/G2^n$	*The percentage of fucosylation (with bisecting GlcNAc) of digalactosylated structures*	*$GP15) / G2^n * 100$*	-0.85 [-1.37:-0.33]	$1.36x10^{-3}$
	FB^n/F^n	*Ratio of fucosylated structures with and without bisecting GlcNAc*	*$FB^n / F^n * 100$*	-0.67 [-1.17:-0.16]	$9.44x10^{-3}$
	$FB^n/F^{n\ total}$	*The incidence of bisecting GlcNAc in all fucosylated structures in total neutral IgG glycans*	*$FB^n / F^{n\ total} * 100$*	-0.69 [-1.19:-0.18]	$7.85x10^{-3}$
	$F^n/(B^n + FB^n)$	*Ratio of fucosylated non-bisecting GlcNAc structures and all structures with bisecting GlcNAc*	*$F^n/(GP13^n + FB^n)$*	0.65 [0.15:1.16]	$1.09x10^{-2}$
	$B^n/(F^n + FB^n)$	*Ratio of structures with bisecting GlcNAc and all fucosylated structures (+/- bisecting GlcNAc)*	*$GP13^n / (F^n + FB^n) * 1000$*	0.18 [-0.31:0.66]	$4.70x10^{-1}$
	$FBG2^n/FG2^n$	*Ratio of fucosylated digalactosylated structures with and without bisecting GlcNAc*	*$GP15^n/GP14^n$*	-0.89 [-1.41:-0.36]	$9.00x10^{-4}$
	$FBG2^n/(FG2^n + FBG2^n)$	*The incidence of bisecting GlcNAc in all fucosylated digalactosylated structures in total neutral IgG glycans*	*$GP15^n/(GP14^n + GP15^n) * 100$*	-0.89 [-1.42:-0.37]	$8.47x10^{-4}$
	$FG2^n/(BG2^n + FBG2^n)$	*Ratio of fucosylated digalactosylated non-bisecting GlcNAc structures and all digalactosylated structures with bisecting GlcNAc*	*$GP14^n/(GP13^n + GP15^n)$*	0.91 [0.38:1.44]	$7.32x10^{-4}$
	$BG2^n/(FG2^n + FBG2^n)$	*Ratio of digalactosylated structures with bisecting GlcNAc and all fucosylated digalactosylated structures (+/- bisecting GlcNAc)*	*$GP15^n/(GP14^n + GP15^n) * 1000$*	-0.93 [-1.46:-0.39]	$6.56x10^{-4}$



Supplementary Figure 1. UPLC analysis of the IgG glycome. An example of a UPLC chromatogram with graphical representation of glycan structures present in each chromatography peak (GP1– GP24).



Supplementary Figure 2. ROC curves illustrating the performance of regularized logistic regression model in predicting disease status for CKD cases and controls in the discovery population.

Supplementary Table 2. Comparison of Fc IgG glycopeptides analysed by nano LC-MS/MS with released IgG glycans analysed by UPLC.

MS glycans (n=96)		ULPC glycans (n=3212)	
Glycan name	**Beta[95%CI]**	**Glycan name**	**Beta[95%CI]**
IgG1_G2FS1	3.27 [-2.28;8.82]	GP18	1.48 [0.89;2.07]
IgG1_G2F	6.23 [-0.59;13.04]	GP14	1.46 [0.85;2.07]
IgG1_G0FNn	-1.46 [-6.98;4.07]	GP6n	-1.39 [-1.98;-0.80]
IgG1_G2Fn	5.20 [-1.39;11.79]	GP14n	1.29 [0.68;1.90]
IgG1 FBS1/FS1	-4.55 [-9.49;0.40]	FBS1/FS1	-1.12 [-1.65;-0.59]
IgG1 FBS1/(FS1+FBS1)	-5.34 [-10.45;-0.23]	FBS1/(FS1+FBS1)	-1.10 [-1.63;-0.57]
IgG1 G2n	3.44 [-3.04;9.92]	G2n	1.20 [0.60;1.80]
IgG1_G0FN	-1.15 [-6.73;4.43]	GP6	-1.14 [-1.71;-0.57]
IgG1 G0n	-0.71 [-6.81;5.38]	G0n	-1.16 [-1.76;-0.56]
IgG1 FGS1/(F+FG+FGS1)	-0.23 [-5.38;4.91]	FGS/(F+FG+FGS)	1.01 [0.46;1.56]
IgG1_G0n	-2.27 [-8.33;3.80]	GP2n	-0.91 [-1.42;-0.40]
IgG1_G0	-1.88 [-7.97;4.21]	GP2	-0.90 [-1.43;-0.38]

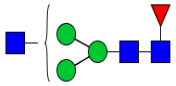
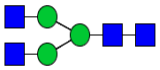
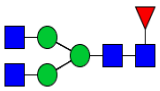
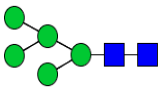
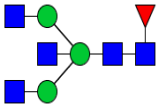
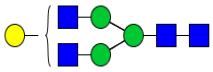
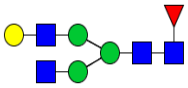
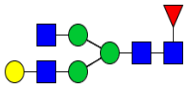
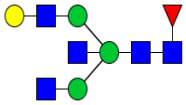
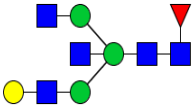
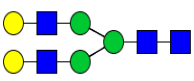
Supplementary Table 3. Association of total plasma glycome and eGFR.

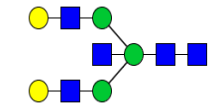
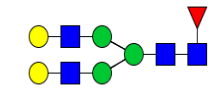
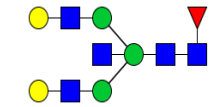
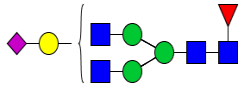
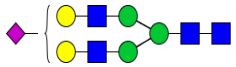
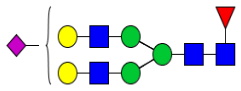
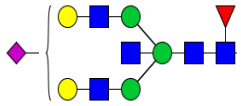
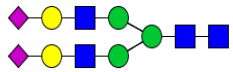
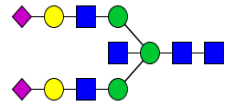
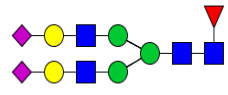
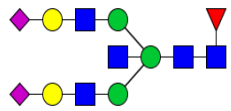
Total plasma glycan peak	*Major glycan(s)*	*B* [95% CI]	p
gly2	M5, FA2B	-1.91 [-3.42:-0.40]	0.01
gly1	FA2	-1.44 [-2.88:-0.00]	0.05
gly42	A4F1G4S4	0.89 [-0.27:2.06]	0.13
gly34	FA3G3S3	0.88 [-0.33:2.09]	0.15
gly41	A4G4S4	0.84 [-0.31:2.00]	0.15
gly6	FA2[6]BG1	-0.96 [-2.32:0.40]	0.17
gly30	A3G3S3	0.76 [-0.53:2.04]	0.25
gly16	A2BG2S1	-0.71 [-2.00:0.58]	0.28
gly40	A4G4S4	0.64 [-0.55:1.84]	0.29
gly4	FA2[6]G1	-0.74 [-2.11:0.63]	0.29
gly39	A4G4S4	0.63 [-0.61:1.87]	0.32
gly13	FA2[3]G1S1	0.60 [-0.58:1.77]	0.32
gly35	A3F1G3S3	0.67 [-0.66:2.00]	0.32
gly33	A3G3S3	0.58 [-0.62:1.78]	0.35
gly10.11	FA2G2	0.76 [-0.92:2.43]	0.38
gly36	A4G4S3	0.55 [-0.71:1.81]	0.39
gly3	A2[6]BG1	-0.53 [-1.82:0.77]	0.43
gly19	FA2BG2S1	-0.53 [-1.84:0.78]	0.43
gly9	A2BG2	-0.49 [-1.70:0.72]	0.43
gly14	FA2[3]G1S1	-0.49 [-1.78:0.81]	0.46
gly7	M6	-0.44 [-1.61:0.73]	0.46

gly38	A4F1G3S3	0.42 [-0.89:1.73]	0.53
gly37	A4G4S3	0.40 [-0.90:1.69]	0.55
gly18	FA2G2S1	0.44 [-1.07:1.94]	0.57
gly26	FA2BG2S2	-0.36 [-1.70:0.99]	0.60
gly24	A2BG2S2	0.28 [-0.99:1.55]	0.67
gly5	FA2[3]G1	-0.29 [-1.64:1.05]	0.67
gly25	FA2G2S2	-0.26 [-1.62:1.11]	0.71
gly29	A3G3S2	0.23 [-1.13:1.59]	0.74
gly31.32	A3G3S3, FA3G3S3	0.20 [-1.01:1.40]	0.75
gly17	M5A1G1S1	-0.19 [-1.52:1.14]	0.78
gly27.28	A3G3S2, A3BG3S2	0.16 [-1.07:1.40]	0.79
gly22	M9	-0.16 [-1.48:1.16]	0.81
gly23	A2G2S2	-0.10 [-1.38:1.17]	0.87
gly12	FA2BG2	-0.07 [-1.46:1.32]	0.92
gly8	A2G2	0.06 [-1.27:1.39]	0.93
gly20.21	A2G2S2	-0.04 [-1.30:1.21]	0.95
gly15	A2G2S1	0.02 [-1.12:1.16]	0.97

8.4. **Supplemental data 4:** The Association Between Low Back Pain and Composition of IgG Glycome

Supplementary Table 1. Structure and description of the studied IgG glycans.

GROUP	Edinburgh Code	Zagreb Code	STRUCTURE[a]	DESCRIPTION	FORMULA
Total IgG glycans (neutral + charged)	IGP1	GP1		*The percentage of FA1 glycan in total IgG glycans*	*GP1 / GP* 100*
	IGP2	GP2		*The percentage of A2 glycan in total IgG glycans*	*GP2 / GP* 100*
	IGP3	GP4		*The percentage of FA2 glycan in total IgG glycans*	*GP4 / GP* 100*
	IGP4	GP5		*The percentage of M5 glycan in total IgG glycans*	*GP5 / GP* 100*
	IGP5	GP6		*The percentage of FA2B glycan in total IgG glycans*	*GP6 / GP* 100*
	IGP6	GP7		*The percentage of A2G1 glycan in total IgG glycans*	*GP7 / GP* 100*
	IGP7	GP8		*The percentage of FA2[6]G1 glycan in total IgG glycans*	*GP8 / GP* 100*
	IGP8	GP9		*The percentage of FA2[3]G1 glycan in total IgG glycans*	*GP9 / GP* 100*
	IGP9	GP10		*The percentage of FA2[6]BG1 glycan in total IgG glycans*	*GP10 / GP* 100*
	IGP10	GP11		*The percentage of FA2[3]BG1 glycan in total IgG glycans*	*GP11 / GP* 100*
	IGP11	GP12		*The percentage of A2G2 glycan in total IgG glycans*	*GP12 / GP* 100*

	IGP12	GP13		*The percentage of A2BG2 glycan in total IgG glycans*	*GP13 / GP* 100*
	IGP13	GP14		*The percentage of FA2G2 glycan in total IgG glycans*	*GP14 / GP* 100*
	IGP14	GP15		*The percentage of FA2BG2 glycan in total IgG glycans*	*GP15 / GP* 100*
	IGP15	GP16		*The percentage of FA2G1S1 glycan in total IgG glycans*	*GP16 / GP * 100*
	IGP16	GP17		*The percentage of A2G2S1 glycan in total IgG glycans*	*GP17/ GP * 100*
	IGP17	GP18		*The percentage of FA2G2S1 glycan in total IgG glycans*	*GP18 / GP * 100*
	IGP18	GP19		*The percentage of FA2BG2S1 glycan in total IgG glycans*	*GP19 / GP * 100*
	IGP19	GP20		*Structure not determined*	*GP20 / GP * 100*
	IGP20	GP21		*The percentage of A2G2S2 glycan in total IgG glycans*	*GP21 / GP * 100*
	IGP21	GP22		*The percentage of A2BG2S2 glycan in total IgG glycans*	*GP22 / GP * 100*
	IGP22	GP23		*The percentage of FA2G2S2 glycan in total IgG glycans*	*GP23 / GP * 100*
	IGP23	GP24		*The percentage of FA2BG2S2 glycan in total IgG glycans*	*GP24 / GP * 100*
Total IgG glycans - derived	IGP24	FGS/(FG+FGS)		*The percentage of sialylation of fucosylated galactosylated structures without*	*SUM(GP16 + GP18 + GP23) / SUM(GP16 + GP18 + GP23 + GP8 + GP9 +*

parameters				*bisecting GlcNAc in total IgG glycans*	*GP14)* 100*
	IGP25	FBGS/(FBG+FBGS)		*The percentage of sialylation of fucosylated galactosylated structures with bisecting GlcNAc in total IgG glycans*	*SUM(GP19 + GP24) / SUM(GP19 + GP24 + GP10 + GP11 + GP15)* 100*
	IGP26	FGS/(F+FG+FGS)		*The percentage of sialylation of all fucosylated structures without bisecting GlcNAc in total IgG glycans*	*SUM(GP16 + GP18 + GP23) / SUM(GP16 + GP18 + GP23 + GP4 + GP8 + GP9 + GP14)* 100*
	IGP27	FBGS/(FB+FBG+FBGS)		*The percentage of sialylation of all fucosylated structures with bisecting GlcNAc in total IgG glycans*	*SUM(GP19 + GP24) / SUM(GP19 + GP24 + GP6 + GP10 + GP11 + GP15)* 100*
	IGP28	FG1S1/(FG1+FG1S1)		*The percentage of monosialylation of fucosylated monogalactosylated structures in total IgG glycans*	*GP16 / SUM(GP16 + GP8 + GP9)* 100*
	IGP29	FG2S1/(FG2+FG2S1+FG2S2)		*The percentage of monosialylation of fucosylated digalactosylated structures in total IgG glycans*	*GP18 / SUM(GP18 + GP14 + GP23)* 100*
	IGP30	FG2S2/(FG2+FG2S1+FG2S2)		*The percentage of disialylation of fucosylated digalactosylated structures in total IgG glycans*	*GP23 / SUM(GP23 + GP14 + GP18)* 100*
	IGP31	FBG2S1/(FBG2+FBG2S1+FBG2S2)		*The percentage of monosialylation of fucosylated digalactosylated structures with bisecting GlcNAc in total IgG glycans*	*GP19 / SUM(GP19 + GP15 + GP24)* 100*
	IGP32	FBG2S2/(FBG2+FBG2S1+FBG2S2)		*The percentage of disialylation of fucosylated digalactosylated structures with bisecting GlcNAc in total IgG glycans*	*GP24 / SUM(GP24 + GP15 + GP19)* 100*
	IGP33	$F^{total}S1/F^{total}S2$		*Ratio of all fucosylated (+/-*	*SUM(GP16 + GP18 + GP19)*

				bisecting GlyNAc) monosyalilated and disialylated structures in total IgG glycans	*/ SUM(GP23 + GP24)*
	IGP34	FS1/FS2		*Ratio of fucosylated (without bisecting GlcNAc) monosialylated and disialylated structures in total IgG glycans*	*SUM(GP16 + GP18) / GP23*
	IGP35	FBS1/FBS2		*Ratio of fucosylated (with bisecting GlcNAc) monosialylated and disialylated structures in total IgG glycans*	*GP19 / GP24*
	IGP36	FBS^{total}/FS^{total}		*Ratio of all fucosylated sialylated structures with and without bisecting GlcNAc*	*SUM(GP19 + GP24) / SUM(GP16 + GP18 + GP23)*
	IGP37	FBS1/FS1		*Ratio of fucosylated monosialylated structures with and without bisecting GlcNAc*	*GP19 / SUM(GP16 + GP18)*
	IGP38	FBS1/(FS1+FBS1)		*The incidence of bisecting GlcNAc in all fucosylated monosialylated structures in total IgG glycans*	*GP19 / SUM(GP16 + GP18 + GP19)*
	IGP39	FBS2/FS2		*Ratio of fucosylated disialylated structures with and without bisecting GlcNAc*	*GP24 / GP23*
	IGP40	FBS2/(FS2+FBS2)		*The incidence of bisecting GlcNAc in all fucosylated disialylated structures in total IgG glycans*	*GP24 / SUM(GP23 + GP24)*
Neutral IgG glycans	IGP41	$GP1^n$		*The percentage of FA1 glycan in total neutral IgG glycans (GP^n)*	*$GP1 / GP^n * 100$*
	IGP42	$GP2^n$		*The percentage of A2 glycan in total neutral IgG glycans (GP^n)*	*$GP2 / GP^n * 100$*

	IGP43	$GP4^n$		*The percentage of FA2 glycan in total neutral IgG glycans (GP^n)*	*$GP4 / GP^n * 100$*
	IGP44	$GP5^n$		*The percentage of M5 glycan in total neutral IgG glycans (GP^n)*	*$GP5 / GP^n * 100$*
	IGP45	$GP6^n$		*The percentage of FA2B glycan in total neutral IgG glycans (GP^n)*	*$GP6 / GP^n * 100$*
	IGP46	$GP7^n$		*The percentage of A2G1 glycan in total Ineutral IgG glycans (GP^n)*	*$GP7 / GP^n * 100$*
	IGP47	$GP8^n$		*The percentage of FA2[6]G1 glycan in total neutral IgG glycans (GP^n)*	*$GP8 / GP^n * 100$*
	IGP48	$GP9^n$		*The percentage of FA2[3]G1 glycan in total neutral IgG glycans (GP^n)*	*$GP9 / GP^n * 100$*
	IGP49	$GP10^n$		*The percentage of FA2[6]BG1 glycan in total neutral IgG glycans (GP^n)*	*$GP10 / GP^n * 100$*
	IGP50	$GP11^n$		*The percentage of FA2[3]BG1 glycan in total neutral IgG glycans (GP^n)*	*$GP11 / GP^n * 100$*
	IGP51	$GP12^n$		*The percentage of A2G2 glycan in total neutral IgG glycans (GP^n)*	*$GP12 / GP^n * 100$*
	IGP52	$GP13^n$		*The percentage of A2BG2 glycan in total neutral IgG glycans (GP^n)*	*$GP13 / GP^n * 100$*
	IGP53	$GP14^n$		*The percentage of FA2G2 glycan in total neutral IgG glycans (GP^n)*	*$GP14 / GP^n * 100$*
	IGP54	$GP15^n$		*The percentage of FA2BG2 glycan in total neutral IgG glycans (GP^n)*	*$GP15 / GP^n * 100$*
Neutral IgG glycans - derived paramet	IGP55	$G0^n$		*The percentage of agalactosylated structures in total neutral IgG glycans*	*SUM($GP1^n$: $GP6^n$)*
	IGP56	$G1^n$		*The percentage of monogalactosylate d structures in total*	*SUM($GP7^n$: $GP11^n$)*

ers				*neutral IgG glycans*	
	IGP57	$G2^n$		*The percentage of digalactosylated structures in total neutral IgG glycans*	*SUM($GP12^n$: $GP15^n$)*
	IGP58	$F^{n\ total}$		*The percentage of all fucosylated (+/- bisecting GlcNAc) structures in total neutral IgG glycans*	*SUM($GP1^n$+ $GP4^n$+ $GP5^n$+ $GP6^n$+ $GP8^n$+ $GP9^n$+ $GP10^n$+ $GP11^n$+ $GP14^n$+ $GP15^n$)*
	IGP59	$FG0^{n\ total}/G0^n$		*The percentage of fucosylation of agalactosylated structures*	*SUM($GP1^n$+ $GP4^n$+ $GP5^n$+ $GP6^n$) / $G0^n$ * 100*
	IGP60	$FG1^{n\ total}/G1^n$		*The percentage of fucosylation of monogalactosylated structures*	*SUM($GP8^n$+ $GP9^n$+ $GP10^n$+ $GP11^n$) / $G1^n$ * 100*
	IGP61	$FG2^{n\ total}/G2^n$		*The percentage of fucosylation of digalactosylated structures*	*SUM($GP14^n$+ GP15) / $G2^n$ * 100*
	IGP62	F^n		*The percentage of fucosylated (without bisecting GlcNAc) structures in total neutral IgG glycans*	*SUM($GP1^n$+ $GP4^n$+ $GP5^n$+ $GP8^n$+ $GP9^n$+ $GP14^n$)*
	IGP63	$FG0^n/G0^n$		*The percentage of fucosylation (without bisecting GlcNAc) of agalactosylated structures*	*SUM($GP1^n$+ $GP4^n$+ $GP5^n$) / $G0^n$ * 100*
	IGP64	$FG1^n/G1^n$		*The percentage of fucosylation (without bisecting GlcNAc) of monogalactosylated structures*	*SUM($GP8^n$+ $GP9^n$) / $G1^n$ * 100*
	IGP65	$FG2^n/G2^n$		*The percentage of fucosylation (without bisecting GlcNAc) of digalactosylated structures*	*$GP14^n$/ $G2^n$ * 100*
	IGP66	FB^n		*The percentage of fucosylated (with bisecting GlcNAc) structures in total neutral IgG glycans*	*SUM($GP6^n$ + $GP10^n$ + $GP11^n$ + $GP15^n$)*
	IGP67	$FBG0^n/G0^n$		*The percentage of fucosylation (with*	*$GP6^n$/ $G0^n$ * 100*

				bisecting GlcNAc) of agalactosylated structures	
	IGP68	$FBG1^n/G1^n$		*The percentage of fucosylation (with bisecting GlcNAc) of monogalactosylated structures*	*$SUM(GP10^n + GP11^n) / G1^n * 100$*
	IGP69	$FBG2^n/G2^n$		*The percentage of fucosylation (with bisecting GlcNAc) of digalactosylated structures*	*$GP15) / G2^n * 100$*
	IGP70	FB^n/F^n		*Ratio of fucosylated structures with and without bisecting GlcNAc*	*$FB^n / F^n * 100$*
	IGP71	$FB^n/F^{n\ total}$		*The incidence of bisecting GlcNAc in all fucosylated structures in total neutral IgG glycans*	*$FB^n / F^{n\ total} * 100$*
	IGP72	$F^n/(B^n + FB^n)$		*Ratio of fucosylated non-bisecting GlcNAc structures and all structures with bisecting GlcNAc*	*$F^n/(GP13^n + FB^n)$*
	IGP73	$B^n/(F^n + FB^n)$		*Ratio of structures with bisecting GlcNAc and all fucosylated structures (+/- bisecting GlcNAc)*	*$GP13^n/(F^n + FB^n) * 1000$*
	IGP74	$FBG2^n/FG2^n$		*Ratio of fucosylated digalactosylated structures with and without bisecting GlcNAc*	*$GP15^n/GP14^n$*
	IGP75	$FBG2^n/(FG2^n + FBG2^n)$		*The incidence of bisecting GlcNAc in all fucosylated digalactosylated structures in total neutral IgG glycans*	*$GP15^n/(GP14^n + GP15^n) * 100$*
	IGP76	$FG2^n/(BG2^n + FBG2^n)$		*Ratio of fucosylated digalactosylated non-bisecting GlcNAc structures and all digalactosylated structures with bisecting GlcNAc*	*$GP14^n/(GP13^n + GP15^n)$*
	IGP77	$BG2^n/(FG2^n + FBG2^n)$		*Ratio of digalactosylated structures with*	*$GP15^n/(GP14^n + GP15^n) * 1000$*

				bisecting GlcNAc and all fucosylated digalactosylated structures (+/- bisecting GlcNAc)	

[a]Symbol nomenclature for glycan representation was used according to: Varki, A. *et al.* Symbol nomenclature for glycan representation. *Proteomics* **9,** 5398–9 (2009).

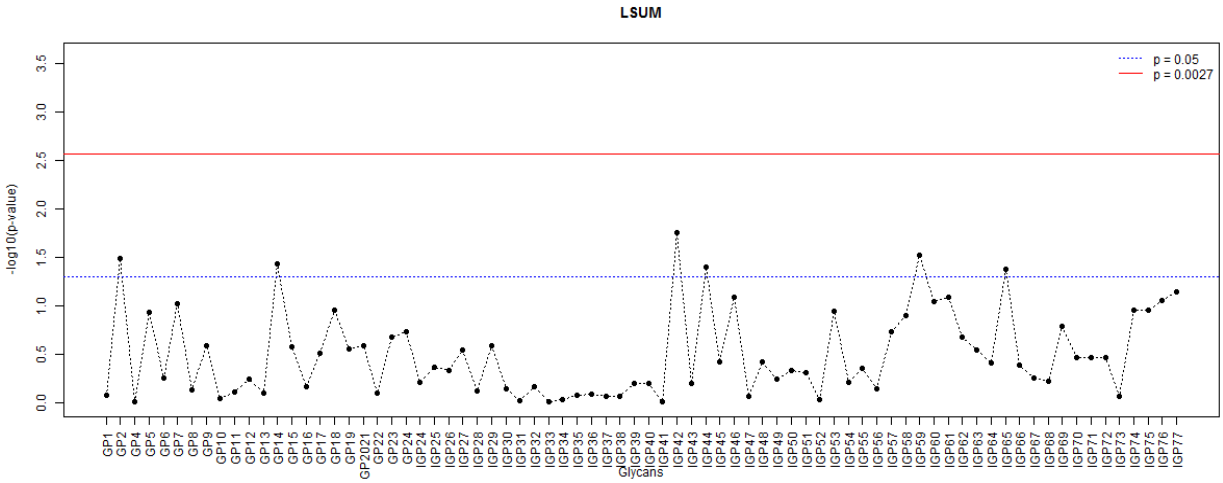
Supplementary Table 2. Correlations between glycan levels and summary scores for lumbar magnetic resonance imaging signs.

Glycan	Correlation coefficient*	p-value
GP1	-0.008	0.835
GP2	0.085	0.032
GP4	-0.001	0.974
GP5	0.063	0.115
GP6	0.024	0.549
GP7	0.066	0.096
GP8	-0.014	0.725
GP9	-0.045	0.255
GP10	0.005	0.907
GP11	0.011	0.781
GP12	0.022	0.574
GP13	-0.011	0.781
GP14	-0.083	0.037
GP15	-0.044	0.265
GP16	-0.016	0.681
GP17	0.041	0.306
GP18	-0.064	0.109
GP19	-0.043	0.281
GP2021	0.045	0.258
GP22	-0.011	0.783
GP23	-0.050	0.211

GP24	-0.053	0.183
IGP24	-0.021	0.606
IGP25	-0.031	0.433
IGP26	-0.029	0.462
IGP27	-0.043	0.281
IGP28	0.012	0.755
IGP29	0.045	0.260
IGP30	0.014	0.720
IGP31	0.003	0.938
IGP32	-0.017	0.670
IGP33	0.002	0.962
IGP34	0.004	0.917
IGP35	0.009	0.831
IGP36	0.009	0.819
IGP37	0.008	0.846
IGP38	0.008	0.847
IGP39	0.019	0.634
IGP40	0.019	0.634
IGP41	0.001	0.976
IGP42	0.095	0.017
IGP43	0.020	0.623
IGP44	0.082	0.040
IGP45	0.035	0.381
IGP46	0.069	0.082
IGP47	0.007	0.864
IGP48	-0.035	0.378
IGP49	0.023	0.569
IGP50	0.030	0.459
IGP51	0.027	0.491
IGP52	0.004	0.917

IGP53	-0.063	0.115
IGP54	-0.021	0.607
IGP55	0.031	0.442
IGP56	0.014	0.718
IGP57	-0.053	0.183
IGP58	-0.061	0.126
IGP59	-0.086	0.030
IGP60	-0.067	0.091
IGP61	-0.070	0.081
IGP62	-0.050	0.211
IGP63	-0.043	0.285
IGP64	-0.035	0.383
IGP65	-0.081	0.042
IGP66	0.033	0.406
IGP67	0.023	0.560
IGP68	0.021	0.594
IGP69	0.056	0.162
IGP70	0.038	0.341
IGP71	0.038	0.341
IGP72	-0.038	0.342
IGP73	0.007	0.851
IGP74	0.064	0.111
IGP75	0.064	0.111
IGP76	-0.068	0.087
IGP77	0.072	0.072

* Pearson's correlations were estimated after adjusting glycan levels for age, sex, BMI, and inflammatory disease status.



Supplementary Figure 1. P-values (-log10) for the analysis of correlations between glycan levels and LSUM MRI scores. P-values correspond to Pearson's correlation coefficients provided in Supplementary table 2.

9. BIOGRAPHY

Toma Keser was born on 25th of January 1988 in Zagreb, Croatia. He finished elementary school in Zaprešić and the Fifth Gymnasium in Zagreb. In 2006 he started studying Pharmacy at the Faculty of Pharmacy and Biochemistry, University of Zagreb. During his studies he was awarded the Rector's Prize for scientific work "*N*-glycosylation of serum proteins and transferrin in the early course of acute pancreatitis". His Master thesis: "Determination of *N*-glycan profile from dried blood spots" was also from the same field. He graduated in 2011. From 2011 till 2012 he worked at Genos Ltd in Glycobiology Laboratory as a researcher - PhD student. Since 2012 he is working as a research and teaching assistant at the Department of Biochemistry and Molecular Biology, at the Faculty of Pharmacy and Biochemistry, where he also started his PhD in Pharmacy. His main field of interest is glycobiology, especially glycosylation of immunoglobulin G and total plasma proteins. In 2013 he visited the Leiden University Medical Center (LUMC) for a research and training visit for six months. During his visit he worked in the Glycomics and Glycoproteomics group at the Center for Proteomics and Metabolomics where he got training in using different mass spectrometric methods for the analysis of protein glycosylation. In 2015 he visited the Estonian Genome Center, University of Tartu for one month, where he prepared samples from Estonian Biobank for glycan analysis. Toma is a member of the Croatian Society of Biochemistry and Molecular Biology. He co-authored 9 peer-reviewed papers and one chapter in a Springer Protocols Handbook and participated in many international scientific conferences.

List of Publications

Freidin, M. B., **Keser, T.**, Gudelj, I., Štambuk, J., Vučenović, D., Allegri, M., Pavić, T., Šimurina, M., Fabiane, S. M., Lauc, G. & Williams, F. M. K. The Association between Low Back Pain and Composition of IgG Glycome. *Sci Rep.* **6**, 26815 (2016).

Barrios, C., Zierer, J., Gudelj, I., Štambuk, J., Ugrina, I., Rodríguez, E., Soler, M. J., Pavić, T., Šimurina, M., **Keser, T.**, Pučić-Baković, M., Mangino, M., Pascual, J., Spector, T. D,, Lauc, G. & Menni, C. Glycosylation Profile of IgG in Moderate Kidney Dysfunction. *J Am Soc Nephrol*. **27,** 933–41 (2016).

Pavić, T., Gudelj, I., **Keser, T.**, Pučić-Baković, M. & Gornik, O. Enrichment of hydrophobic membrane proteins using Triton X-114 and subsequent analysis of their N-glycosylation. *Biochim Biophys Acta.* (2015) [Epub ahead of print].

Vučković, F., Krištić, J., Gudelj, I., Teruel, M., **Keser, T.**, Pezer, M., Pučić-Baković, M., Štambuk, J., Trbojević-Akmačić, I., Barrios, C., Pavić, T., Menni, C., Wang, Y., Zhou, Y., Cui, L., Song, H., Zeng, Q., Guo, X., Pons-Estel, B. A., McKeigue, P., Leslie Patrick, A., Gornik, O., Spector, T. D., Harjaček, M., Alarcon-Riquelme, M., Molokhia, M., Wang, W. & Lauc, G. Association of systemic lupus erythematosus with decreased immunosuppressive potential of the IgG glycome. *Arthritis Rheumatol*. **67,** 2978–89 (2015).

Gudelj, I., **Keser, T.**, Vučković, F., Škaro, V., Goreta, S. Š., Pavić, T., Dumić, J., Primorac, D., Lauc, G. & Gornik, O. Estimation of human age using N-glycan profiles from bloodstains. *Int J Legal Med*. **129,** 955–61 (2015).

Krištić, J., Vučković, F., Menni, C., Klarić, L., **Keser, T.**, Beceheli, I., Pučić-Baković, M., Novokmet, M., Mangino, M., Thaqi, K., Rudan, P., Novokmet, N., Sarac, J., Missoni, S., Kolčić, I., Polašek, O., Rudan, I., Campbell, H., Hayward, C., Aulchenko, Y., Valdes, A., Wilson, J. F., Gornik, O., Primorac, D., Zoldoš, V., Spector, T. & Lauc, G. Glycans are a novel biomarker of chronological and biological ages. *J Gerontol A Biol Sci Med Sci.* **69,** 779–89 (2014).

Novokmet, M., Lukić, E., Vučković, F., Đurić, Ž., **Keser, T.**, Rajšl, K., Remondini, D., Castellani, G., Gašparović, H., Gornik, O. & Lauc, G. Changes in IgG and total plasma protein glycomes in acute systemic inflammation. *Sci Rep.* **4,** 4347 (2014).

Lauc, G., Huffman, J. E., Pučić, M., Zgaga, L., Adamczyk, B., Mužinić, A., Novokmet, M., Polašek, O., Gornik, O., Krištić, J., **Keser, T.**, Vitart, V., Scheijen, B., Uh, H. W., Molokhia, M., Patrick, A. L., McKeigue, P., Kolčić, I., Lukić, I. K., Swann, O., van Leeuwen, F. N., Ruhaak, L. R., Houwing-Duistermaat, J. J., Slagboom, P. E., Beekman, M., de Craen, A. J., Deelder, A. M., Zeng, Q., Wang, W., Hastie, N. D., Gyllensten, U., Wilson, J. F., Wuhrer, M., Wright, A. F., Rudd, P. M., Hayward, C., Aulchenko, Y., Campbell, H. & Rudan, I. Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. *PLoS Genet.* **9,** e1003225 (2013).

Menni, C., **Keser, T.**, Mangino, M., Bell, J. T., Erte, I., Akmačić, I., Vučković, F., Pučić-Baković, M., Gornik, O., McCarthy, M. I., Zoldoš, V., Spector, T. D., Lauc, G. & Valdes, A. M. Glycosylation of immunoglobulin G: role of genetic and epigenetic influences. *PLoS One*. **8,** e82558 (2013).

Basic documentation card

University of Zagreb
Faculty of Pharmacy and Biochemistry
Department of Biochemistry and Molecular Biology
A. Kovačića 1, 10000 Zagreb, Croatia

PhD thesis

INFLUENCE OF GENETIC AND ENVIRONMENTAL FACTORS ON *N*-GLYCOSYLATION OF IMMUNOGLOBULIN G AND TOTAL PLASMA PROTEINS DETERMINED BY TWIN STUDY

Toma Keser

SUMMARY

Glycans constitute the most abundant and diverse form of the post-translational modifications. While genes unequivocally determine the structure of each polypeptide, there is no genetic template for the glycan part. Instead, hundreds of genes and their products interact in the very complex pathway of glycan biosynthesis which is further complicated by environmental influences. Therefore, the aim of this thesis was to determine the extent to which individual differences in immunoglobulin G and total plasma proteins glycosylation patterns reflect genetic versus environmental influences. A twin study design was used and study subjects were twins enrolled in the TwinsUK registry, a national register of adult twins. More than 4500 samples were analyzed by HILIC-UPLC (Hydrophilic Interaction Ultra Performance Liquid Chromatography). A high contribution of the genetic component to *N*-glycome composition was found. Variation in levels of 51 of the 76 IgG glycan traits studied was at least 50% heritable and only a small proportion of *N*-glycan traits had a low genetic contribution. Heritability of plasma *N*-glycome was also high, with half of the plasma glycan traits being at least 50% heritable. Further, epigenome-wide association (EWA) analysis showed that methylation levels at some genes are also implicated in glycome composition, both in those with high heritability and those with a lower genetic contribution. The study to investigate the potential role of glycosylation in kidney function was also conducted. Fourteen IgG glycan traits were associated with renal function in discovery population and remained significant after validation in an independent subset of monozygotic twins discordant for renal disease, reflecting difference in galactosylation, sialylation, and level of bisecting *N*-acetylglucosamine. Using the weighted correlation network analysis (WGCNA) for IgG glycan traits, a correlation between low back pain (LBP) and glycan modules was established. There was a weak positive correlation between pain phenotypes and "pro-antibody-dependent cell-mediated cytotoxicity (ADCC)" WGCNA glycan modules (high bisecting *N*-acetylglucosamine and low core fucose) and a weak negative correlation between pain phenotypes and "anti-ADCC" module (high core fucose, no bisecting *N*-acetylglucosamine). This suggests that glycans are promising candidates for biomarkers in many different diseases.

The thesis is deposited in the Central Library of Faculty of Pharmacy and Biochemistry.

Thesis includes: 95 pages, 15 figures, 15 tables and 122 references. Original is in English language.

Keywords: IgG glycome, total plasma glycome, *N*-glycosylation, glycan analysis, HILIC, twin study, heritability, chronic kidney disease, low back pain

Menthor: **Gordan Lauc**, **Ph.D.** *Full Professor, Faculty of Pharmacy and Biochemistry, University of Zagreb.*

Reviewers: **Olga Gornik**, **Ph.D.** *Associate Professor, Faculty of Pharmacy and Biochemistry, University of Zagreb.*

Jerka Dumić, **Ph.D.** *Full Professor, Faculty of Pharmacy and Biochemistry, University of Zagreb.*

Ingeborg Barišić, **Ph.D.** *Full Professor, School of Medicine, University of Zagreb.*

The thesis accepted: July 2016.

Temeljna dokumentacijska kartica

Sveučilište u Zagrebu **Doktorski rad**
Farmaceutsko-biokemijski fakultet
Zavod za biokemiju i molekularnu biologiju
A. Kovačića 1, 10000 Zagreb, Hrvatska

UTJECAJ GENSKIH I OKOLIŠNIH ČIMBENIKA NA *N*-GLIKOZILACIJU IMUNOGLOBULINA G I UKUPNIH PLAZMATSKIH PROTEINA ODREĐEN STUDIJOM NA BLIZANCIMA

Toma Keser

SAŽETAK

Glikani predstavljaju najzastupljeniji i najraznolikiji oblik posttranslacijske modifikacije. Dok geni nedvosmisleno određuju strukturu svakog polipeptida, za sintezu glikana ne postoji genski predložak. Umjesto toga, stotine gena i njihovih produkata sudjeluju u vrlo kompleksnoj biosintezi glikana koju okolišni utjecaji čine još složenijom. Stoga je cilj ovog doktorskog rada odrediti razmjer kojim genski i okolišni čimbenici utječu na *N*-glikane imunoglobulina G (IgG) i ukupnih glikoproteina plazme. Da bi se postigao navedeni cilj, upotrijebljena je studija na blizancima. Ispitanici su regrutirani TwinsUK registrom, najvećim registrom blizanaca u Velikoj Britaniji. Više od 4500 uzoraka analizirano je HILIC-UPLC metodom (kromatografijom vrlo visoke djelotvornosti temeljenoj na hidrofilnim interakcijama, eng. *Hydrophilic Interaction Ultra Performance Liquid Chromatography*). Velika genska komponenta (heritabilnost ≥ 50%) pokazana je za 51 od 76 glikanskih svojstava IgG-a. Nasuprot tome, samo je 12 glikanskih svojstava IgG-a pokazalo malu gensku komponentu. Heritabilnost plazmatskog *N*-glikoma također se pokazala velikom. Polovica plazmatskih glikanskih svojstava bila je barem 50% heritabilna. Epigenomska asocijacijska analiza pokazala je da razina metilacije na nekim genima također utječe na sastav glikoma, i to na visokoheritabilne i niskoheritabilne glikane. Također je provedeno prvo istraživanje potencijalne uloge IgG glikozilacije u funkciji bubrega. U prvoj je analizi pronađena značajna povezanost s bubrežnom funkcijom za 14 glikanskih svojstava, a ostala je značajna i nakon validacije na neovisnoj podskupini monozigotnih blizanaca diskordantnih za bubrežnu funkciju. Ta glikanska svojstva pripadaju trima glavnim glikozilacijskim karakteristikama IgG-a: galaktozilaciji, sijalinizaciji i razini račvajućeg N-acetilglukozamina. Koristeći se WGCNA (eng. *weighted correlation network analysis*) metodologijom, provedena je mrežna anliza razina IgG glikana kod blizanaca, da bi se uspostavili klasteri koreliranih glikana. Pronađene su povezanosti između tih klastera i fenotipova boli kod blizanaca s LBP-om. Opažena je pozitivna korelacija između fenotipova boli i „pro-stanična citotoksičnost ovisna o protutijelima (eng. *antibody-dependent cell-mediated cytotoxicity*, ADCC)“ WGCNA glikanskog modula (visoka razina račvajućeg *N*-acetilglukozamina i niska razina sržne fukoze) te negativna korelacija između fenotipova boli i „anti-ADCC“ modula (visoka razina sržne fukoze, bez račvajućeg *N*-acetilglukozamina). Rezultati pokazuju da su glikani obećavajući biomarkeri za mnoge bolesti.

Rad je pohranjen u Centralnoj knjižnici Farmaceutsko-biokemijskog fakulteta Sveučilišta u Zagrebu.

Rad sadrži: 95 stranica, 15 grafičkih prikaza, 15 tablica i 122 literaturna navoda. Izvornik je na engleskom jeziku.

Ključne riječi: IgG glikom, ukupni plazmatski glikom, *N*-glikozilacija, analiza glikana, HILIC, studija na blizancima, kronična bubrežna bolest, bol u donjem dijelu leđa

Mentor: **Dr.sc. Gordan Lauc**, *redoviti profesor, Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu.*

Ocjenjivači: **Dr.sc. Olga Gornik**, *izvanredni profesor, Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu.*

Dr.sc. Jerka Dumić, *redoviti profesor, Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu.*

Dr.sc. Ingeborg Barišić, *redoviti profesor, Medicinski fakultet Sveučilišta u Zagrebu.*

Rad prihvaćen: srpanj 2016.