


RESEARCH

Open Access



The genetics and epidemiology of N- and O-immunoglobulin A glycomics

Alessia Visconti^{1,2†}, Niccolò Rossi^{1†}, Albert Bondt³, Agnes Hipgrave Ederveen³, Gaurav Thareja⁴, Carolien A. M. Koeleman³, Nisha Stephan⁴, Anna Halama⁴, Hannah J. Lomax-Browne⁵, Matthew C. Pickering⁵, Xu-jie Zhou^{6,7,8,9}, Manfred Wuhrer^{3†}, Karsten Suhre^{4†} and Mario Falchi^{1*†} 

Abstract

Background Immunoglobulin (Ig) glycosylation modulates the immune response and plays a critical role in ageing and diseases. Studies have mainly focused on IgG glycosylation, and little is known about the genetics and epidemiology of IgA glycosylation.

Methods We generated, using a novel liquid chromatography-mass spectrometry method, the first large-scale IgA glycomics dataset in serum from 2423 twins, encompassing 71 N- and O-glycan species.

Results We showed that, despite the lack of a direct genetic template, glycosylation is highly heritable, and that glycopeptide structures are sex-specific, and undergo substantial changes with ageing. We observe extensive correlations between the IgA and IgG glycomes, and, exploiting the twin design, show that they are predominantly influenced by shared genetic factors. A genome-wide association study identified eight loci associated with both the IgA and IgG glycomes (*ST6GAL1*, *ELL2*, *B4GALT1*, *ABCF2*, *TMEM121*, *SLC38A10*, *SMARCB1*, and *MGAT3*) and two novel loci specifically modulating IgA O-glycosylation (*C1GALT1* and *ST3GAL1*). Validation of our findings in an independent cohort of 320 individuals from Qatar showed that the underlying genetic architecture is conserved across ancestries.

Conclusions Our study delineates the genetic landscape of IgA glycosylation and provides novel potential functional links with the aetiology of complex immune diseases, including genetic factors involved in IgA nephropathy risk.

Keywords Immunoglobulin A, Glycosylation, IgA/IgG shared genetics

[†]Alessia Visconti and Niccolò Rossi contributed equally to this work.

[†]Manfred Wuhrer, Karsten Suhre, and Mario Falchi share senior authorship.

*Correspondence:

Mario Falchi
mario.falchi@kcl.ac.uk

¹ Department of Twin Research and Genetic Epidemiology, King's College London, London, UK

² Center for Biostatistics, Epidemiology and Public Health, Department of Clinical and Biological Sciences, University of Turin, Turin, Italy

³ Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands

⁴ Department of Biophysics and Physiology, Weill Cornell Medicine–Qatar, Doha, Qatar

⁵ Centre for Inflammatory Disease, Department of Immunology and Inflammation, Imperial College London, London, UK

⁶ Renal Division, Peking University First Hospital, Beijing, China

⁷ Peking University Institute of Nephrology, Beijing, China

⁸ Key Laboratory of Renal Disease, Ministry of Health of China, Beijing, China

⁹ Key Laboratory of Chronic Kidney Disease Prevention and Treatment, Ministry of Education, Peking University, Beijing, China



Background

Immunoglobulins are large molecules produced by plasma cells and activated B cells in response to exposure to antigens. Immunoglobulins (Ig) undergo complex co- and post-translational modifications including the selective addition of oligosaccharides (glycans) to the side chains of asparagine (*N*-glycosylation), serine or threonine (*O*-glycosylation) residues in the peptide backbone. These modifications, despite being non-template driven, are genetically determined [1, 2]. So far, IgG has attracted the interest of most of the epidemiological and genetic research in Ig glycobiology, with studies showing that IgG glycosylation affects effector functions [3] with relevant implications in ageing and disease, including, among others, autoimmune and infectious diseases, and cancer [3]. Conversely, less is known about the genetics and epidemiology of the IgA glycome. IgA is the most abundantly secreted immunoglobulin in the body, the second most prevalent antibody isotype in blood after IgG, and is heavily glycosylated [1–4]. Notably, while the CH2 domain of each of the four IgG subclasses (IgG_{1–4}) contains a single *N*-glycosylation site [5], the two IgA subclasses (IgA_{1–2}) carry multiple conserved *N*-glycosylation sites [5]. In addition, IgA₁ can also undergo *O*-glycosylation [4]. Aberrant *O*-glycosylation has been observed in cancer and autoimmunity [6], suggesting a role for both IgA *O*- and *N*-glycosylation in health and disease. Abnormally glycosylated IgA₁ has been found to be the driving pathogenic force in IgA nephropathy, the most common form of glomerulonephritis worldwide. Here, we explore the genetic architecture of the circulating IgA *O*- and *N*-glycome and compare glycosylation features between the IgA and IgG molecules to assess the impact of shared genetics and environment on these differently specialised antibody isotypes.

Methods

Discovery cohort

The study subjects were 2423 individuals of self-declared European ancestry from the TwinsUK cohort [7] (Additional file 1: Table S1). TwinsUK includes about 14,000 monozygotic and dizygotic twin volunteers from all regions across the UK, unselected for any disease and representative of the general population.

Replication cohort

The Qatar Metabolomics Study on Diabetes (QMDiab) is a cross-sectional case–control study of diabetes with 374 participants of Arab, South Asian, and Filipino descent [8, 9] (Additional file 1: Table S1). For 320 samples, joint genetics and glycomics data were available. All study participants were enrolled between February 2012 and June

2012 at the Dermatology Department of Hamad Medical Corporation in Doha, Qatar.

Mass spectrometric glycomics

In both twinsUK and QMDiab, IgA and IgG glycans were quantified in serum samples using liquid chromatography–mass spectrometry (LC–MS), as previously described [10, 11].

In TwinsUK, blood samples were collected from participants between 1996 and 2016 and were centrifuged at 2000 g for 10 min at room temperature, and instantly stored at –80 °C before processing. Plasma samples were shipped by courier on dry ice to the Center for Proteomics and Metabolomics, Leiden University Medical Center, and 10 µL of plasma was used for IgG and IgA analysis.

In QMDiab, non-fasting blood samples were collected according with standard protocols as previously described [8]. Briefly, blood for IgA analysis was collected using EDTA tubes, and centrifuged at 2500 g for 10 min at room temperature, plasma was collected, aliquoted and stored at –80 °C until analysis. For each sample, 2 µL and 5 µL of plasma were used for IgG and IgA analysis, respectively.

In total, 33 IgA₁ *O*-linked glycopeptides, 38 IgA_{1–2} *N*-linked glycopeptides, and 36 IgG_{1–4} *N*-linked glycopeptides were retained and quantified. The absolute signal intensities were normalised to the intensity sum of all glycopeptide species sharing the same tryptic peptide sequence, resulting in relative abundances. In this manuscript, IgA₁ and IgA₂ glycopeptide names are composed of the letter codes of the first three amino acids of the peptide sequence: HYT, LSL, TPL, SES, ENI, and LAG, the last detected in two variants (*i.e.*, as LAGC and LAGY). The peptide name is followed by the glycan composition indicating the number of hexoses (H), *N*-acetylhexosamines (N), fucoses (F), and sialic acids (Fig. 1a, Additional file 1: Table S2). Glycan structures were assigned on the basis of the tandem mass spectrometric analyses of glycopeptides performed with method development [12] and IgA-released glycan analysis supported by spiking experiments with human plasma standards [13].

Structurally similar glycopeptides were summarised into derived traits calculated from their relative intensities (Additional file 1: Table S3), resulting in 7 IgA₁ *O*-glycan, 21 IgA_{1–2} *N*-glycan, and 16 IgG_{1–4} *N*-glycan derived traits. Because each measured glycopeptide structure carries different types of monosaccharides, derived traits can give a more composite and robust measure of the different glycosylation features. Outliers, defined as measurements deviating more than four standard deviations from the mean of each trait, were removed. To ensure the

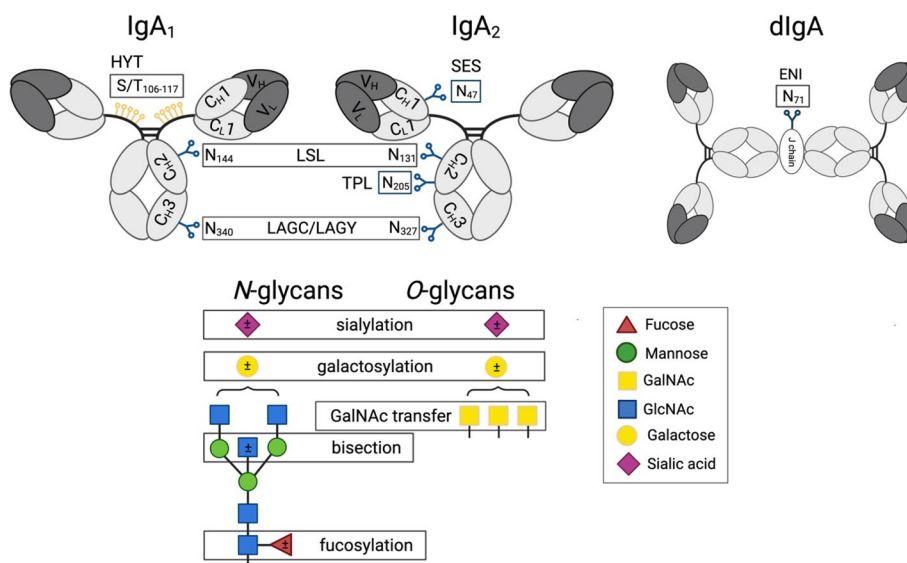


Fig. 1 Schematic representation of IgA₁ and IgA₂ with examples for O- and N-glycan structures. N-glycans are attached to the amino side chain of an asparagine residue (N), while O-glycans are attached to the side chain of serine (S) or threonine (T). Each IgA₁ heavy chain contains two N-glycosylation sites (i.e., at N144 and N340), while the hinge region has six O-glycosylation sites (i.e., at T106, T109, S111, S113, T114, and T117), all present on a single tryptic peptide. Both dimeric IgA₁ and IgA₂ have one additional N-glycosylation site at the J-chain (i.e., at N71). The three-letter code defines the tryptic (glyco-)peptides: HYT for the O-glycosylation at the hinge region of IgA₁; LSL for the N-glycosylated sites N144 or N131 on IgA₁ or IgA₂, respectively; LAG for the N-glycosylated sites N340 or N327 on IgA₁ or IgA₂, respectively, which were detected with either a terminal tyrosine (LAGY) or as the truncated form (LAGC); SES for the N-glycosylated site N47 on IgA₂; TPL for the N-glycosylated site N205 on IgA₂; ENI for the N-glycosylated site N71 at the J-chain on IgA₂. Glycosylation site numbering is according to Uni-ProtKB. Monosaccharide symbols and example structures of O- and N-glycans and of derived traits are also shown

normality of their distribution, measured glycopeptides and derived traits were quantile normalised.

Association with age and sex

The association of measured glycopeptides and derived traits with age at sample collection and sex was assessed by fitting a linear mixed model in the R statistical framework (*lmer* function, *lme4* package [14] v1.1–31) with plate number and family structure modelled as random effects. We used the approach proposed by Li and Ji [15] to determine the effective number of independent tests to control for the family-wise error rate, resulting in 37 and 24 tests, for IgA and IgG, respectively. Associations were considered significant if the p-value was lower than $0.05/37 = 1.35 \times 10^{-3}$ and $0.05/24 = 2.08 \times 10^{-3}$ for IgA and IgG, respectively.

Heritability estimation in TwinsUK

We used the *mets* R package [16] (v1.2.5) to estimate the contribution of additive genetic (A), shared (C) and individual-specific environment (E) effects on measured glycopeptides and derived traits variations (ACE model), using a classical twin design study. The ACE model was then compared with the most parsimonious AE model, which did not include the effect of the common

environment, and the CE and E model, which hypothesised that the trait variability only under environmental control. Models were compared through Akaike’s information criterion (AIC). Quantile-normalised measured glycopeptides and derived traits were corrected for plate number (included as a random effect in a linear mixed model, *lmer* R function), and passed to *mets* . Sex and age at the sample collection were included as covariates. Four pairs of monozygotic twins and three pairs of dizygotic twins were removed because the twins were either adopted or reared apart.

IgA and IgG intra-class correlations

For both isotypes, intra-class (i.e., IgA-IgA and IgG-IgG) pairwise correlation between both measured glycopeptides and derived traits was estimated using Pearson’s product–moment correlation on pre-processed glycan abundances. More in detail, quantile-normalised measured glycopeptides and derived traits were adjusted for age, sex (fixed effects), plate number, and family structure (random effects) using a linear mixed model (*lmer* R function), and the correlations were computed on the residuals. Associations were considered significant if the absolute correlation coefficient was greater than 0.25 and the p-value was lower than $0.05/(N \times (N - 1)/2)$, where

the denominator corresponds to the number of elements of a lower triangular matrix with size $N \times N$, and N corresponds to the effective number of independent tests as described above. The value of N was 37 and 24, for IgA and IgG, respectively.

IgA glycan ratios

Ratios were calculated using untransformed relative frequencies of pairs of measured glycopeptides expressed on the same peptide, after outliers were removed. For N -glycans, we selected IgA and IgG measured glycopeptides representing the product–substrate of known sialylation and galactosylation reactions in the Ig N -glycosylation pathway [17]. For O -glycans, reactions of N -acetylgalactosaminylation, sialylation, and galactosylation were estimated based on the ratios of pairs of measured O -glycopeptides which differed for a single GalNAc, sialic acid and galactose residue, respectively. Overall, 21 multi-step enzymatic conversions were identified, involving at least two enzymatic transformations of the same underlying glycan structure, as exemplified below:



Differences in the distribution of ratios representing two consecutive glycosylation reactions were assessed using the Wilcoxon test, and we considered significant p -values lower than $0.05/31 = 2.38 \times 10^{-3}$, where 31 corresponds to the number of comparisons performed.

IgA-IgG shared heritability

Pairwise additive genetic (ρ_G) and environmental (ρ_E) correlations among IgA-IgG derived traits were estimated through bivariate maximum likelihood-based variance decomposition in SOLAR-Eclipse [18] (v8.1.1; <http://www.solar-eclipse-genetics.org/>). Age and sex were included as covariates. SOLAR-Eclipse was also used to estimate the phenotypic correlations (ρ_P) so that the same proportion of variance associated with covariates is removed for phenotypic and genetic correlation estimates. Likelihood ratio tests were used to assess whether ρ_P , ρ_G , and ρ_E were significantly different from zero. Phenotypic correlations were considered significant if the absolute correlation coefficient was larger than 0.25 and the associated p -value was lower than $0.05/(37 \times 24) = 5.63 \times 10^{-5}$, where 37 and 24 correspond, respectively, to the effective number of independent IgA and IgG, estimated as described above.

For pairs of correlated IgA-IgG glycan traits, we estimated the genetic (COV_G) and environmental (COV_E) covariances based on the following formulas:

$$COV_G(x, y) = r_G(x, y) \times \sqrt{h_x^2} \times \sqrt{h_y^2}$$

$$COV_E(x, y) = r_E(x, y) \times \sqrt{1 - h_x^2} \times \sqrt{1 - h_y^2}$$

where h_x^2 and h_y^2 are the heritabilities (h^2) of traits x and y , as estimated by SOLAR-Eclipse, and $r_G(x, y)$ and $r_E(x, y)$ are the genetic and environmental correlations, respectively. Genetic/environmental correlation coefficients were constrained to zero if the associated p -value was larger than 0.05, so that the resulting covariance is zero. Then, the proportion of traits phenotypic correlation explained by genetics (*i.e.*, shared heritability) was estimated based on the following equation:

$$h_{xy}^2 = \frac{COV_G(x, y)}{COV_G(x, y) + COV_E(x, y)}$$

which assumes that the total phenotypic covariance is given by the sum of the genetic and environmental covariances COV_G and COV_E , respectively.

Genotyping

In TwinsUK, microarray genotyping was conducted using a combination of Illumina arrays (HumanHap300, HumanHap610, 1 M-Duo and 1.2 M-Duo 1 M) and imputation was performed using the Michigan Imputation Server [19] using haplotype information from the Haplotype Reference Consortium (HRC) panel [20] (r1.1). Genotypes were available for 2371 individuals with glycomics data. A total of 5,411,460 single nucleotide polymorphisms (SNPs) meeting the following conditions were included in our genome-wide association study: call rate $\geq 95\%$, minor allele frequency (MAF) $> 5\%$, and imputation score > 0.4 .

In QMDiab, genotyping was carried out using Illumina Omni array 2.5 (v8). Standard quality control of genotyped data was applied, with SNPs filtered by sample call rate $> 98\%$; SNP call rate $> 98\%$, and Hardy–Weinberg equilibrium (HWE) $P < 1 \times 10^{-6}$. Imputation was done using SHAPEIT software [21] with 1000G phase 3 version 5 and mapped to the GRCh37 human genome build. Imputed SNPs were filtered by imputation quality > 0.7 .

Genome-wide association study (GWAS) discovery step

To take into account the non-independence of the twin data, the association with IgA and IgG measured glycopeptides and derived traits in the TwinsUK cohort was performed using GEMMA [22] (v0.98.1), assuming an additive genetic model and including sex, age at the sample collection, the first five principal components assessed on the genomic data, and plate number as covariates. Associations were considered significant

and taken forward for replication if their discovery p -value was lower than $5 \times 10^{-8}/37 = 1.35 \times 10^{-9}$ and $5 \times 10^{-8}/24 = 2.08 \times 10^{-9}$, for IgA and IgG, respectively.

Identification of independent signals within loci

We used a stepwise procedure to identify independent signals within the loci identified in the discovery cohort. For each locus, we fitted a new regression model, where the top-associated genome-wide significant SNP was included as a covariate (i.e., conditional model). Genome-wide significant ($P < 5 \times 10^{-8}$) SNP resulting from the conditional model was considered as an independent signal and was included in the covariate set of a new conditional model. We stopped the stepwise procedure when we could not identify any additional genome-wide significant SNPs. Conditional models were built using GEMMA [22] (v0.98.1).

Replication step

We attempted replication of independent signals for IgA measured glycopeptides and derived traits using data from the QMDiab cohort. Replication was evaluated at the lead SNP of each locus, or at a tag SNP in high linkage disequilibrium ($R^2 \geq 0.8$, distance ≤ 500 kb) with any genome-wide significant SNP within the locus. The linkage disequilibrium (LD) structure was assessed with LDlink [23] (v3.6) using the available GBR population from 1000 genomes project phase 3. Genetic association was conducted using a linear regression model adjusting for age, sex, diabetes status, and the first three genetic principal components, and was considered replicated if the direction of the effects was concordant between the two cohorts, and if the p -value was lower than $0.05/72 = 6.94 \times 10^{-3}$, where 72 is the number of IgA measured glycopeptides and derived traits-genomic locus pair identified in the discovery step.

To assess the replicability of loci reported by previous GWAS for serum IgG *N*-glycans, we retrieved the reported lead SNPs for the 27 independent genome-wide significant loci identified by the largest ($n = 8,090$) GWAS published to date on circulating IgG glycome [2]. We considered an association replicated in TwinsUK if the p -value of the association for any IgG measured glycopeptides and derived traits at the lead SNP of each locus was lower than $0.05/(27 \times 24) = 7.72 \times 10^{-5}$, where 27 is the number of loci tested and 24 is the effective number of IgG measured glycopeptides and derived traits tested.

Identification of known loci

We interrogated the NHGRI-EBI GWAS Catalog [24] (release 2023-02-03, association $P < 5 \times 10^{-8}$) to identify previously reported associations in coincidence or

in strong linkage disequilibrium with those identified and replicated in our study. Additionally, we queried our results to identify associations in coincidence or in strong linkage disequilibrium with 25 genome-wide significant loci reported by a large IgAN meta-analysis [25], and passing a Bonferroni-derived threshold of $0.05/25 = 0.002$. Linkage disequilibrium was assessed using LDlink [23] with the parameters listed above.

Variants annotation

Locus-gene mapping was performed using OpenTarget [26] (<https://genetics.opentargets.org/>, release 22.10, October 2022). OpenTarget used the “locus-to-gene” (L2G) model to prioritise likely causal genes. An L2G score is derived for a given SNP–gene pair by aggregating evidence from different mapping strategies, including physical distance from the transcription start site of nearby genes, (sQTL) and expression (eQTL) quantitative trait loci colocalization, and chromatin interaction mapping, as well as in silico predictions of SNP functional consequence.

Shared genetic effect

For each pair of derived traits for which a replicated genome-wide significant signal was identified, we applied a Bayesian bivariate analysis, as implemented in the GWAS-PW software [27] to investigate the presence of a shared genetic effect in the TwinsUK cohort. Briefly, GWAS-PW estimates the posterior probability of quasi-independent genomic regions to include a genetic variant which (i) associated only with the first derived trait (model 1), (ii) associated only with the second derived trait (model 2), (iii) associated with both derived traits (model 3) or (iv) that the genomic block includes two genetic variants, associating independently with each of the two derived traits (model 4). We defined regions for which a genome-wide significant signal was identified for at least one of the two traits and showed a posterior probability larger than 0.9 for model 3 or model 4 as characterised by an underlying shared genetic effect or by colocalization of effects, respectively.

Quasi-independent regions ($n = 1703$) were pre-determined by estimating LD blocks in European populations [16]. Since our GWASs were performed on overlapping sets of individuals, we estimated the expected correlation in the effect sizes from the GWAS summary statistics using *fgwas* [28], as suggested in [27], and used this correction factor in the GWAS-PW modelling.

SNP-based heritability

For a given SNP_{*i*}, we calculated the proportion of explained phenotypic variance in TwinsUK as:

$$\sigma_i = 2 \times p_i \times q_i \times \beta_i^2$$

where β_i is the estimated effect of SNP_{*i*} in univariate analysis, and p_i and q_i are the minor and major allele frequencies of SNP_{*i*} respectively, as estimated in TwinsUK. For measured glycopeptides and derived traits that had more than one independent genome-wide significant association, we estimated the total SNP-based variance explained as the sum of the proportion of variance explained by each of the top independent SNPs.

GWAS signals contribution to IgA-IgG shared heritability

To evaluate the contribution of GWAS signals to IgG-IgA glycome shared heritability, we estimated pairwise IgA-IgG derived traits additive genetic covariance upon conditioning on the lead SNPs in TwinsUK. More in detail, for each pair tested, derived traits were independently adjusted for age, sex, plate number, and lead SNPs at genome-wide-significant loci associated with any of the two traits (*lmer* R function). Standardised residuals were then passed to SOLAR-Eclipse [18] (v8.1.1) to estimate the conditional phenotypic and genetic correlations. The additive genetic covariance was estimated based on the following formula:

$$\text{Cov}_G(x, y) = A_x \times A_y \times \rho G_{xy}$$

where A_x and A_y are the square roots of the heritabilities (h^2) of traits x and y , and ρG_{xy} is the additive genetic correlation.

Results

Cohort description

Study subjects were 2423 individuals of European ancestry from the TwinsUK cohort [7] (513 monozygotic and 615 dizygotic twin pairs, and 167 singletons), mainly females (78%), between 19 and 88 (mean=58, standard deviation (SD)=11) years old (Additional file 1: Table S1). Using LC-MS, we quantified 33 O-linked and 38 N-linked IgA₁₋₂ glycan species at the glycopeptide level [10], whose names are here composed of the letter codes of the first three amino acids of the tryptic peptide sequence (Fig. 1a, Additional file 1: Table S2). Additionally, we summarised glycosylation features (e.g., bisection, fucosylation, galactosylation, sialylation) across structurally related O- and N- glycans expressed on the same glycopeptide into 7 and 21 derived traits, respectively (Additional file 1: Table S3). In the same sample, we also quantified 36 IgG₁₋₄ N-linked glycans and calculated 16 IgG₁₋₄ derived traits [9] (Additional file 1: Tables S2 and S3). Derived traits provide a composite and more robust measure of Ig glycosylation than single measured glycopeptides [29] and will be the main focus of this work.

IgA glycosylation is sex-specific and undergoes substantial changes with ageing

We found that IgA N-glycan bisection, regardless of the glycopeptide on which it was expressed, was higher in females ($P \leq 4.53 \times 10^{-9}$; Additional file 1: Tables S4 and S5). Age was significantly associated with 68% of the IgA-derived traits (Additional file 1: Tables S4 and S5), indicating increased N-glycan bisection and decreased N- and O-glycan sialylation and galactosylation with ageing. IgG glycosylation data showed concordant results, consistent with previous reports [30] (Additional file 1: Tables S4 and S5).

IgA glycosylation is highly heritable

Using the classical twin model, we showed IgA glycosylation to be significantly heritable, with derived traits displaying mean $h^2=50.8\%$ (SD=13.8%; max $h^2=74.3\%$; Fig. 2), similarly to IgG derived traits (mean $h^2=52.2\%$, SD=12.1%, max $h^2=71.0\%$; Additional file 2: Fig. S1, Additional file 1: Tables S6 and S7). Bisectioned N-glycans were among the highest heritable traits in both isotypes. A shared environmental component was estimated to contribute (mean $c^2=21.3\%$; SD=2.2%) to the variance of sialylation and galactosylation of diantennary IgA N-glycans at the LSL and SES glycopeptides (Fig. 2, Additional file 1: Tables S6 and S7). Interestingly, IgA measured O-glycans where overall more influenced by environmental factors than N-glycans (Wilcoxon $P=3.29 \times 10^{-4}$).

IgA and IgG glycomes show extensive genetic correlation

After Bonferroni-correction, and at $|\rho|>0.25$, we observed an extensive IgA and IgG measured glycopeptides intra-class pairwise correlation (Additional file 1: Tables S8 and S9), particularly for those expressed on the same glycopeptide, as expected, since most of them represented substrates and products of enzymatic reactions. Interestingly, using their ratios as proxies for enzymatic reaction rates, we observed a consistent reduction in enzymatic efficiency along with growing substrate complexity for progressive N-acetylgalactosaminylation, sialylation, and galactosylation of IgA₁ O-glycan structures ($P < 2.2 \times 10^{-16}$; Fig. 3a; Additional file 2: Fig. S2 and S4). A progressive reduction in enzymatic efficiency with growing glycan chains was systematically observed across different IgA and IgG N-glycan molecules (Additional file 2: Fig. S3 and S5), mainly due to stochastic effects associated with decreasing number of sites available for glycosylation, with decreased substrate availability potentially also playing a role.

After Bonferroni-correction, and at $|\rho|>0.25$, we observed also extensive intra-class pairwise correlation among IgA and IgG-derived traits (Fig. 3c, Additional

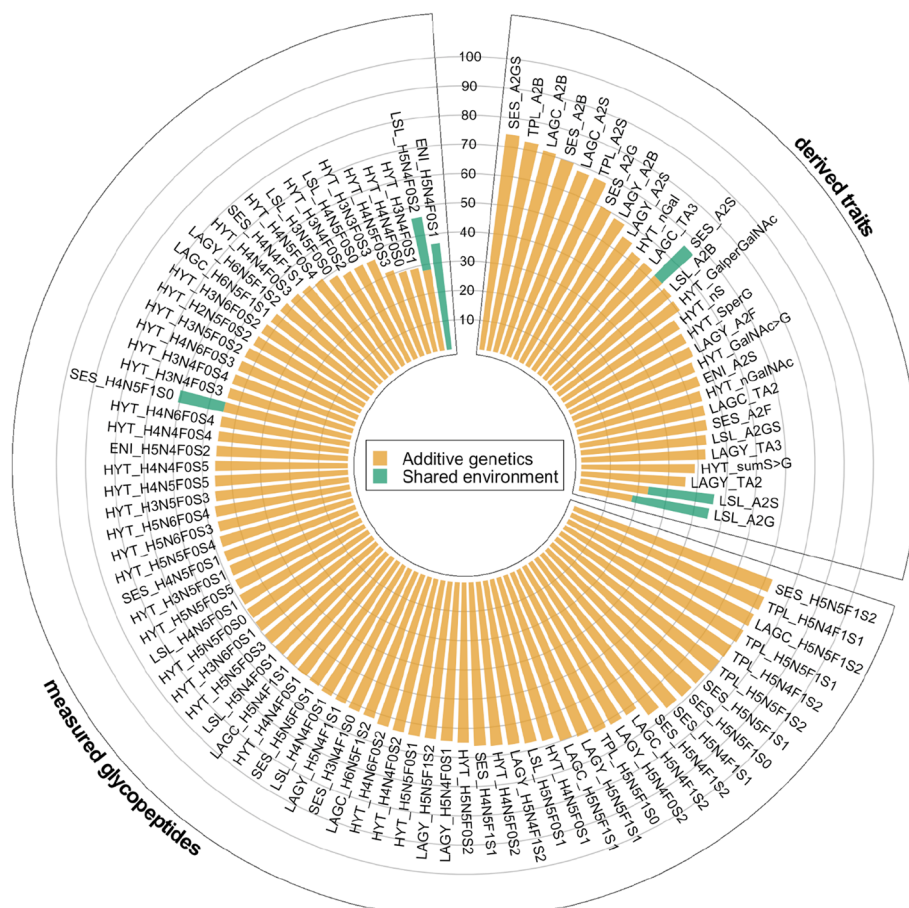


Fig. 2 Heritability of IgA measured glycopeptides and derived traits. We used the ACE model to partition the variance for each trait into additive genetic (orange), and shared (green) and unique (white) environmental components

file 1: Tables S10 and S11), particularly between those involving the same glycosylation linkage (*N*- or *O*-, Additional file 1: Table S12). More interestingly, at a Bonferroni-corrected threshold of $P < 5.63 \times 10^{-5}$ and at $|\rho| > 0.25$, we observed 69 inter-class pairwise correlations between IgA and IgG-derived traits (Additional file 1: Table S13), albeit with weaker strength compared to intra-class IgA and IgG correlations (Additional file 2: Fig. S6), thus suggesting that shared genetic and/or environmental factors can simultaneously influence the behaviour of different antibody isotypes (Additional file 1: Table S10). Using a bivariate variance component model, we showed these correlations between IgA and IgG-derived traits to be mostly explained by shared genetic factors (median shared heritability = 79%, IQR = 73–83%; Additional file 1: Table S14), with shared environmental factors also playing a significant role (median contribution 21%; IQR = 17–27%). The largest shared genetic component was observed for those pairs of IgA-IgG derived traits describing *N*-glycans sialylation and bisection (Fig. 3b), the latter also

being the most heritable trait in both IgA and IgG.

Genetics of the IgA glycome

To identify genes involved in this shared heritability and explore the genetic architecture of the serum IgA glycome, we performed a genome-wide association study (GWAS) of IgA glycosylation in 2371 individuals at 5,411,460 common (minor allele frequency, MAF > 5%) single nucleotide polymorphism (SNPs), using a linear mixed model to account for non-independence of observations within twin pairs, and adjusting for sex, age, LC-MS plate number, and the first five principal components of the genotype data. The average genomic inflation factor was 1.01 (max = 1.04), thus suggesting that there was no residual confounding by population stratification nor cryptic relatedness, nor any apparent systematic genotyping error. At a Bonferroni-corrected threshold of $P < 1.35 \times 10^{-9}$ (Methods), we identified 87 genome-wide significant associations of 45 measured glycopeptides and 17 derived traits at 11 genetic loci (Fig. 4, Table 1, Additional file 1: Table S15, Additional

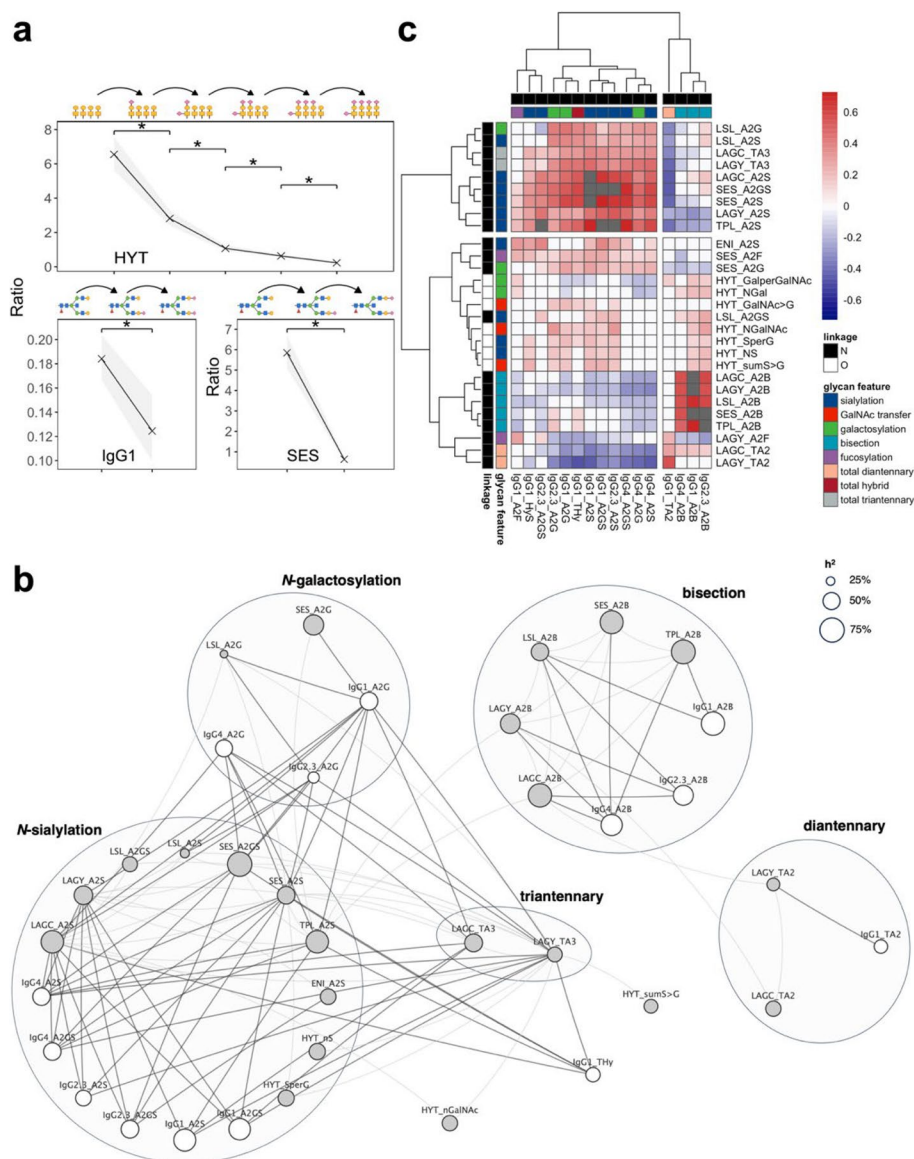


Fig. 3 **a** Progressive decrease in sialylation efficiency of different *N*- and *O*-glycan structures. For each enzymatic reaction depicted on top of the panel, the conversion rate was estimated as the product/substrate ratio using untransformed glycans relative frequencies. For each ratio, the median value is represented with a cross at the corresponding positions of the x-axis, while the grey area shows the interquartile range. Within each panel, glycan structures differ by a single sialic acid residue, and thus reflect sequential sialylation reactions in the glycosylation pathway. Significant differences, evaluated by means of the Wilcoxon test, are indicated with an asterisk ($P < 2.2 \times 10^{-16}$). **b** Correlation network of IgA and IgG *N*-linked glycosylation derived traits. Each node represents a derived trait (IgG: white; IgA: grey), with size proportional to the trait heritability. Edges connect phenotypically correlated (Pearson's $|\rho| > 0.25$) derived traits. IgA-IgG correlations are shown with darker, thicker lines compared IgA-IgA correlations. For IgA, only correlations between derived traits on different peptides are shown. Derived traits are grouped according to their glycosylation features. **c** Genetic correlation of IgA and IgG-derived traits. The heatmap shows IgA-derived traits in rows and IgG-derived traits in columns, with colour labels indicating the linkage (*O*- or *N*-) and the glycosylation feature (e.g., sialylation, fucosylation, bisection). Cells colours represent the genetic correlation coefficients with grey cells indicating IgA-IgG pairs whose calculation did not converge. Derived traits are hierarchically clustered based on absolute genetic correlation coefficients using the *hclust* function as implemented in the *pheatmap* R package (v1.0.12)

file 3). Conditional association analyses identified no further independent genome-wide significant associations. Lead SNPs explained, on average, 5% of the phenotypic

variance of the associated IgA traits (IQR=3–6%, calculated on measured glycopeptides and derived traits) and 9% of their genetic variance (IQR=6–13%; Additional

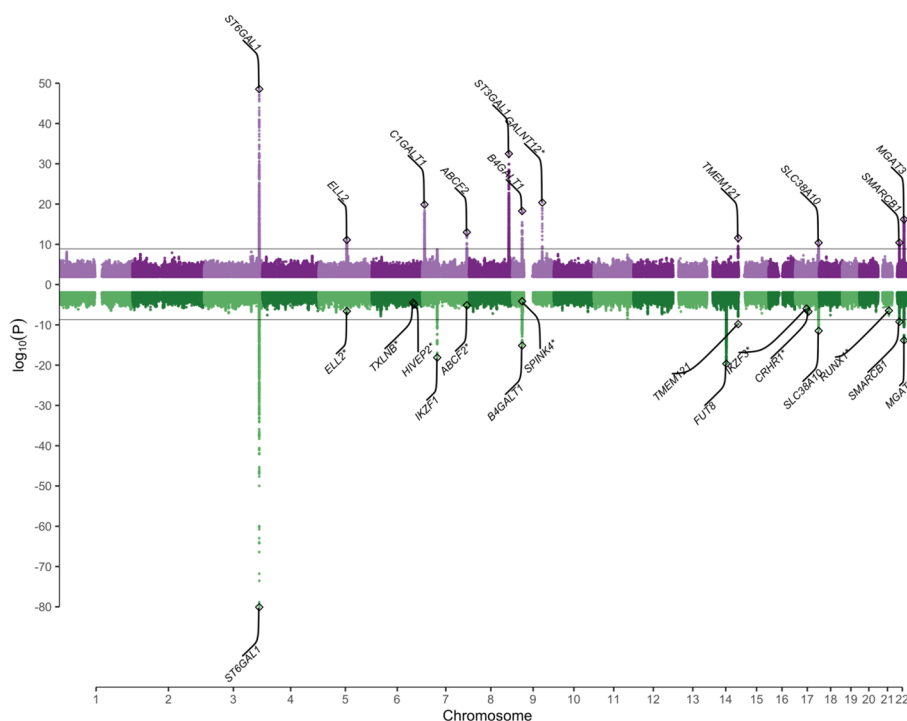


Fig. 4 Miami plot showing the genome-wide association of IgA (top) and IgG measured glycopeptides and derived traits (bottom). The x-axis shows the genomic coordinates (GRCh37.p13) of the tested SNPs and the y-axis shows the $-\log_{10} P$ value of their association. The horizontal black line indicates the threshold for genome-wide significance at 1.35×10^{-9} and 2.08×10^{-9} , for IgA and IgG, respectively. Asterisks indicate newly identified IgA glycosylation loci failing replication in QMDiab (top panel) and replicated IgG loci not reaching genome-wide significance in the present study (bottom panel)

file 1: Tables S16 and S17). Novel genetic associations for O-glycan traits were identified at the genes encoding for the glycosylation enzymes core 1 synthase, glycoprotein-N-acetylgalactosamine 3-β-galactosyltransferase (C1GALT1) and ST3 β-Galactoside α-2,3-Sialyltransferase 1 (ST3GAL1), and were replicated in an independent multi-ancestry cohort of 320 individuals from Qatar (QMDiab [8, 9]; Additional file 1: Table S1) at a conservative Bonferroni-corrected threshold of $P < 6.94 \times 10^{-4}$ (Methods; Table 1 and Additional file 1: Table S15). The lead SNP rs10246303-T at the C1GALT1 locus tagged reported associations for decreased modified Tiffeneau-Pinelli index [31], with an association for chronic obstructive pulmonary disease (COPD) being reported within the same locus [32] (Table 1, Additional file 1: Table S15). A third newly identified locus for O-glycans mapping to the gene encoding for the glycosylation enzyme polypeptide N-Acetylgalactosaminyltransferase 12 (GALNT12) showed comparable effects between TwinsUK and QMDiab, but did not reach statistical significance in the replication, likely due to the low frequency (MAF < 1.5%) of the associated SNPs in QMDiab (MAF_{TwinsUK} > 7%; Additional file 1: Table S15).

At the remaining eight loci identified, associations for O-glycans (ABCF2 and SLC38A10) and N-glycans (ABCF2, B4GALT1, ELL2, MGAT3, ST6GAL1, SMARCB1, and TMEM121) were in strong linkage disequilibrium (LD) ($r^2 > 0.8$) with GWAS associations previously reported for N-glycan traits [2] (Table 1 and Additional file 1: Table S15). IgA N-glycans associations at the ST6GAL1, MGAT3, and ELL2 loci were also identified in QMDiab (Table 1 and Additional file 1: Table S15), with lead SNPs at the ST6GAL1 and ELL2 loci being coincident or in strong linkage disequilibrium (LD) with previously reported associations for IgA nephropathy and multiple myeloma risk, respectively (Table 1).

Shared variants influence both IgA and IgG glycosylation

GWAS of the IgG N-glycan traits replicated previous associations ($P < 7.72 \times 10^{-5}$; Methods; Fig. 4; Additional file 1: Table S18) at 16 out of the 27 loci identified by the largest ($n = 8,090$) GWAS published to date using an ultra-high-performance liquid chromatography (UHPLC)-measured IgG N-glycome [2], with eight loci reaching genome-wide significance ($P < 2.08 \times 10^{-9}$; Additional file 1: Tables S18 and S19, Additional file 4), thus

Table 1 Results of the IgA GWAS. For each locus, the table reports the cytogenic location, the candidate gene along with its function, the total number of associated glycan traits (either measured glycopeptides or derived trait), as well as the fraction, expressed as a percentage, of replicated glycan associations in QMDiab (% Replicated), the strongest associated SNP, along with the effect allele (EA, the minor allele), the other allele (OA), and the effect allele frequency (EAF) in TwinsUK. For each lead SNP within the locus, we report the glycan trait showing the strongest association, along with its GWAS summary statistics, i.e., the association effect size (β), standard error (SE) and p -value (P) in both TwinsUK and QMDiab. We further report whether the locus is associated with either N- or O-glycans in TwinsUK (N/O glycan), as well as information on previous GWAS (identified as [PMID]) reporting associations within the locus for the N-glycome (Glycosylation measurements) or relevant traits or disease (Phenotypic trait/Disease), as extracted from the GWAS Catalog

Locus coordinates	Mapped gene	Function	N. associated Glycans	Top glycan trait	Lead SNP	EA	OA	EAF	β_{TwinsUK}	SE _{TwinsUK}	P _{TwinsUK}	N/O-glyc	Glycosylation measurements [PMID]	Phenotypic trait/disease [PMID]	β_{QMDiab}	SE _{QMDiab}	P _{QMDiab}	% Replicated
3:186705790–186782999	ST6GAL1	Sialyl-trans-ferase	20	LAGC_A2S	rs6764279	A	C	28.4	-0.489	0.032	2.90×10^{-49}	N	IgG [29535710, 23382691, 28878392, 35332118, 32128391] Total plasma [31163085]	IgA nephropathy [26028593]	-0.459	0.073	8.80×10^{-10}	42
5:95221337–95310985	ELL2	Transcription regulator	4	TPL_ _{H5NSF1S2}	rs11744881	T	A	25.3	-0.241	0.035	8.10×10^{-12}	N	IgG [28878392]	Multiple myeloma [26007630, 27363682]	-0.314	0.086	3.10×10^{-4}	25
7:7194680–7290732	CTGALT1	Galactosyltransferase	10	HYT_ _{H3NSF0S2}	rs10246303	T	A	40.4	-0.297	0.032	1.30×10^{-20}	O		Serum galactose-deficient IgA1 levels in IgA nephropathy [33593824]; Lung function (FEV1/FVC) [30804560, 28166213]	-0.388	0.071	1.05×10^{-7}	75
7:150906453–150954562	ABCF2	Trans-porter	3	HYT_ _{H3NSF0S1}	rs113810201	G	A	11.5	-0.376	0.050	1.00×10^{-13}	N/O	IgG [23382691, 32128391, 35332118]	Multiple myeloma [27363682, 33875642]	-0.072	0.148	6.29×10^{-1}	0
8:134509215–134535985	ST3GAL1	Sialyl-trans-ferase	23	HYT_ _{H4NSF0S1}	rs6995270	C	T	49.9	-0.396	0.032	3.20×10^{-33}	O			-0.507	0.068	1.20×10^{-12}	89
9:33119241–33180813	B4GALT1	Galactosyltransferase	6	TPL_ _{H5N4F1S2}	rs3780480	T	G	25.3	-0.315	0.035	5.50×10^{-19}	N	IgG [29535710, 23382691, 28878392, 32128391] Total plasma [31163085]		-0.172	0.078	2.95×10^{-2}	0
9:101364062–101570336	GALNT12	N-acetyl-galactosaminyltransferase	5	HYT_ _{H4N6F0S2}	rs1137654	T	A	9.3	-0.551	0.058	4.30×10^{-21}	O			-0.451	0.323	1.64×10^{-1}	0

Table 1 (continued)

Locus coordinates	Mapped gene	Function	N. associated Glycans	Top glycan trait	Lead SNP	EA	OA	EAF	β_{twinsUK}	SE_{twinsUK}	P_{twinsUK}	N/O-glyc	Glycosylation measurements [PMID]	Phenotypic trait/disease [PMID]	β_{GMDiab}	SE_{GMDiab}	P_{GMDiab}	% Replicated
14:105966605–106019014	<i>TMEM121</i>	Trans-membrane protein	2	SES_H4NSF1S1	rs61983272	C	G	33.1	0.298	0.042	2×10^{-12}	N	IgG [35332118, 28878392, 32128391]					NA
17:79217478–79268562	<i>SLC38A10-CEP131-TEP5IN-NDUFAF8</i>		5	HYT_H4N6F0S4	rs2659005	T	C	43.4	0.206	0.031	4.00×10^{-11}	O	IgG [32128391, 23382691]		0.076	0.075	3.10×10^{-1}	0
22:24135043–24179922	<i>SMARCB1</i>	Chromatin remodeling	4	TPL_H5N4F1S1	rs9624334	C	G	13.5	0.301	0.045	3.80×10^{-11}	N	IgG [29535710, 23382691, 28878392, 35332118, 32128391] Total plasma [31163085]		0.210	0.120	8.10×10^{-2}	0
22:39775156–39874314	<i>MGAT3</i>	N-acetylglucosaminyltransferase	5	LAGC_A2B	rs5757682	T	G	25.0	0.293	0.035	6.60×10^{-17}	N	IgG [29535710, 23382691, 32128391, 28878392, 35332118]	Factor VIII or von Willebrand factor [35285134, 30386737]	0.294	0.079	2.29×10^{-4}	100

suggesting good cross-platform reproducibility. Lead SNPs explained, on average, 5% of the phenotypic variance of the associated IgG measured glycopeptides and derived traits (IQR=2%–6%, calculated on measured glycopeptides and derived traits) and 8% of their genetic variance (IQR=4%–10%; Additional file 1: Tables S16 and S17).

Our GWAS highlighted coincident associations between IgA and IgG glycosylation at eight loci, and 23 out of the 69 pairs of matching IgA/IgG-derived *N*-glycan traits under significant shared genetic control showed coincident association at the *ST6GAL1*, *B4GALT1*, *MGAT3*, and *SMARCB1* genes. The Bayesian bivariate approach implemented in GWAS-PW [27] suggested that the same genetic variants were influencing both IgA and IgG correlated traits (posterior probability of model 3 > 0.90; Additional file 1: Table S20). Overall, all the genetic variants identified for the 69 genetically correlated IgA-IgG pairs explained only a minor fraction of the estimated genetic correlation (median=6%, IQR=0.4%–10%; Additional file 1: Table S21), thus suggesting that a much larger sample size may be needed and/or that the shared genetic effects cannot be fully captured by common SNPs.

Discussion

IgA is the most abundantly secreted immunoglobulin in the body, and it is heavily glycosylated [4]. Yet, little is known about its genetic architecture, and the shared genetic and environmental factors which shape both the IgA and IgG molecules. Here, we used here a high-throughput LC–MS method to generate the first large-scale dataset for the simultaneous analysis of IgA and IgG glycopeptides in a large cohort of twins from the UK.

We showed that bisected *N*-glycans have a higher abundance in females, possibly reflecting a sex-specific modulation of the immune response, with increased bisecting *N*-acetylglucosamine abundances of IgG in females (identified also in this study) previously observed from childhood [33]. Additionally, we showed that ageing is associated with increased *N*-glycan bisection, and decreased *N*- and *O*-glycan sialylation and galactosylation, consistent with previous findings on IgG [30]. Increased abundances of bisection and, mostly, reduced sialylation and galactosylation on IgG *N*-linked glycopeptides have been connected to several inflammatory and autoimmune conditions, and to ageing [3].

Exploiting the twin design, we identified extensive genetic correlations between the IgA and IgG glycomes, with bisected *N*-glycans being the most heritable traits in both IgA and IgG. The largest shared genetic component was observed for those pairs of IgA-IgG derived traits describing *N*-glycans sialylation and bisection, the latter also being the most heritable derived trait in both IgA

and IgG. Bisection is known to be catalysed by the *Mgat3* enzyme across all *N*-glycans, and the observed strong correlation between IgG and IgA bisection suggests it is performed by plasma cells in a concerted manner. This has already been confirmed in vitro: cytokines and TLR ligands influence IgG bisection [34], and we speculate this can extend to IgA-producing plasma cells. Analogously, the *ST6Gal1* enzyme appears responsible for both IgG and IgA *N*-glycan sialylation, suggesting again a shared regulatory mechanism in IgA- and IgG-producing plasma cells [2, 10]. As expected, we also observe a significant correlation between sialylation and galactosylation traits of IgG, reflecting sialylation dependency on galactosylation. Finally, we identified only weak correlations between *N*- and *O*-glycosylation, reflecting their very different enzymatic regulation.

While IgA-IgG correlations were mostly explained by genetics, shared environmental factors had a non-negligible role, in line with previous observations. Indeed, similar age- and disease-associated IgG and IgA glycosylation signatures have already been observed [12, 35, 36], with parasitic infections, rural vs urban environments, stress, medications use, nutrition, and socio-economic status shaping IgG glycosylation patterns [35, 37]. Some of these environmental exposures are likely to similarly influence the conserved glycosylation machinery in IgA- and IgG-producing plasma cell subclasses, resulting in correlated glycosylation patterns.

Our genome-wide association of the IgA glycome identified ten associated genetic loci, including two novel loci involved in *O*-glycosylation (*C1GALT1* and *ST3GAL1*) and eight loci affecting the abundance of both IgA and IgG glycans (*ST6GAL1*, *ELL2*, *B4GALT1*, *ABCF2*, *TMEM121*, *SLC38A10*, *SMARCB1*, and *MGAT3*). These results were validated using 320 individuals of Arab, South Asian, and Filipino descent, showing that the underlying genetic architecture is conserved across ancestries. Quantifications at the ENI, LAGC/LAGY, and LSL peptide clusters assess the relative abundance of *N*-glycan traits present in both the IgA₁ and IgA₂ subclasses. Similarly, glycan forms for IgG₂ and IgG₃ are captured together by our methodology. While the observed associations between glycosylation genes (*MGAT3*, *ST6GAL1*, and *B4GALT1*) and the expected relevant glycan traits at these peptide clusters are unlikely to be affected by potential differences in the ratio of IgG and IgA subclasses, we cannot exclude the possibility, based on current knowledge, that associations at *ELL2* and *TMEM121* are influenced by different subclass abundances expressing subclass-specific glycosylation patterns.

Five loci identified by our GWAS on the IgA glycome, although previously associated with the IgG glycome, did

not map to genes known to be directly involved in glycan synthesis pathways. The *SMARCB1* and *ELL2* genes are involved in transcription regulation, with the former acting as a core component of the SWI/SNF chromatin remodelling complex and the latter functioning as a transcription elongation factor. *SMARCB1* has been suggested to modulate the expression of the bisecting enzyme Mgat3 [2], while *ELL2* influences secretory-specific Ig production by influencing the expression and splicing of a substantial number of genes in the transition from B cells to antibody-secreting cells, including the Ig heavy chain and those in *N*-glycan biosynthesis [38]. The link between the IgA glycome and the cytosolic ATP-binding cassette transporter ABCF2 and the transmembrane protein TMEM121 remains elusive, as the specific functions of ABCF2 and TMEM121 are yet to be uncovered. Finally, a locus on chromosome 17 nearby the *SLC38A10-CEP131-TEPSIN-NDUFAF8* genes could not be unambiguously mapped, as none of these genes have a function directly linked to immunoglobulin glycosylation.

Coincident associations at the loci encompassing the *C1GALT1* and *ST6GAL1* genes contribute to a better understanding of the factors involved in IgA nephropathy (IgAN) risk. Galactose deficiency of IgA₁ *O*-glycans in the hinge region (gd-IgA) has been repeatedly observed in IgAN [39], with serum gd-IgA being traditionally assessed using ELISA. Association for higher serum gd-IgA levels have been reported at rs10238682-G [40], a positive regulator of *C1GALT1* expression, genome-wide significant in this study and in strong LD ($R^2 > 0.8$) with our lead SNP at the *C1GALT1* gene. *C1GALT1* catalyses the transfer of galactose to *N*-acetylgalactosamine *O*-linked to the hydroxy group of threonine or serine residues. We find here that SNPs at this locus associate with increased galactosylation (HYT_nGal and HYT_GalperGalNAc), and lower proportion of degalactosylated acetylgalactosamine relative to galactose (HYT_nGalNAc > nG) in the hinge region. In a previous high-resolution glycomics case-control study of IgA in IgAN susceptibility carried out by our team, we found that these three derived traits were strongly correlated with ELISA measurements of gd-IgA, but they were not associated with IgAN risk or kidney function [10]. Indeed, we suggested that *N*- and *O*-sialylation, rather than *O*-galactosylation, were likely involved in the pathophysiology of IgAN. In this context, the *ST6GAL1* gene, which encodes a sialyltransferase that adds sialic acid to the terminal galactose of *N*-glycoproteins, has been associated with IgAN risk in 8313 cases and 19,680 controls from China [41], as well as with disease severity and progression [42]. The associated SNP rs7634389 ($r^2 > 0.8$ with our lead SNPs at *ST6GAL1*) associates with decreased

ST6Gal1 expression, and, here, with lower *N*-sialylation at most derived traits (LAGC_A2S, LAGY_A2S, LSL_A2GS, SES_A2GS, SES_A2S, TPL_A2S), overlapping glycan traits associating with IgAN in our previous study and confirming the involvement of *N*-sialylation in disease risk. Additionally, we validated, at a suggestive Bonferroni-derived threshold of $0.05/25 = 0.002$, three loci reported by a recent large (10,146 cases and 28,751 controls) multi-ancestry GWAS of IgAN [25], including the risk allele rs10896045-A at the *OVOLI/RELA* gene associating in our study to decreased galactosylation and consequently decreased sialylation at *O*-glycans ($P = 8.7 \times 10^{-4}$ at HYT_nS > nG; Additional file 1: Table S22).

Associations for decreased modified Tiffeneau-Pinelli index [31] and COPD [32] were reported at the *C1GALT1* locus. IgA is the most prevalent Ig in the lungs and is altered in chronic respiratory disease accompanied by a decreased pulmonary capacity, including COPD [43]. However, little is known about the similarity between IgA glycosylation in the lung mucosa and in serum, with considerable differences already been observed between circulating and salivary IgA glycosylation [11].

Associations for increased serum IgA levels [44] and lower multiple myeloma risk [45, 46] have been reported at rs3815768-T and rs56219066-C, respectively, in LD ($R^2 > 0.8$) with our lead SNP rs11744881-T at *ELL2*, here associated with lower IgA *N*-glycan sialylation. Increased sialylation is found in IgG paraproteins from patients with multiple myeloma [47], where it correlates with metastatic behaviour [48]. As *ELL2* influences both secretory-specific Ig production and *N*-glycan biosynthesis [49], it is suggested to increase the risk of malignant transformation by reducing Ig levels, which may result in slower antigen clearance and prolonged B-cell stimulation [50], possibly in a context of aberrantly glycosylated immunoglobulins. Intriguingly, while *ELL2* associates also with IgG sialylation [1], the multiple myeloma risk allele rs56219066-C shows a stronger association with circulating IgA levels compared to IgG [46], suggesting that IgA, more than IgG, *N*-glycan sialylation may play a role in multiple myeloma.

Conclusions

Here, we show here that serum IgA *N*- and *O*-glycosylation is largely genetically determined, and that the underlying genetic architecture is conserved across ancestries. Ageing is accompanied by increased bisection and decreased *N*- and *O*-glycan sialylation and galactosylation, with bisection being also the most heritable trait. We report profound parallels between IgA and IgG *N*-glycosylation, suggesting shared genetic and environmental mechanisms of immune regulation to

simultaneously orchestrate different specialised antibody isotypes. We identified common genetic mechanisms of immune regulation between IgA and IgG N-glycans. However, the major fraction of the shared IgA-IgG glycome heritability remains unexplained, and the shared environmental determinants remain to be explored. Taken together, these findings expand our understanding of the architecture of circulating Ig glycosylation and provide a potential functional link with the aetiology of complex diseases, including of factors involved in IgAN risk.

Abbreviations

eQTL	Expression quantitative trait loci
GWAS	Genome-wide association study
HWE	Hardy-Weinberg equilibrium
Ig	Immunoglobulin
L2G	Locus-to-gene
LC-MS	Liquid chromatography-mass spectrometry
LD	Linkage disequilibrium
QMDiab	Qatar Metabolomics Study on Diabetes
MAF	Minor allele frequency
SNP	Single nucleotide polymorphisms
sQTL	Splicing quantitative trait loci

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-024-01369-6>.

Additional file 1. Supplementary Tables.

Additional file 2. Supplementary Figures.

Additional file 3. IgA genome-wide significant hits.

Additional file 4. IgG genome-wide significant hits.

Acknowledgements

The authors acknowledge the use of the research computing facility at King's College London, Rosalind (<https://rosalind.kcl.ac.uk>) and King's Computational Research, Engineering and Technology Environment (CREATE, <https://docs.er.kcl.ac.uk/>). We thank the staff of the HMC dermatology department and of WCM-Q for their contribution to QMDiab. We would like to thank Serenella Medici from the University of Sassari for the useful discussions. Finally, we are grateful to all study participants for their invaluable contributions to this study.

Authors' contributions

MF, MW, and KS conceptualised and designed the project. MW devised and supervised the mass spectrometric glycomics. AB performed the mass spectrometric glycomics with assistance from AHE. CAMK performed the mass spectrometric glycomics on the QMDiab cohort. AV and NR carried out the statistical analyses with contributions from MF. NS and AH were responsible for the QMDiab data collection and management. GT performed the imputation of the QMDiab genotypes. KS performed the GWAS replication in the QMDiab cohort. AV, NR, KS, MW, and MF interpreted the results with contributions from HJLB and MCP. AV, NR, and MF wrote the manuscript with contributions from XZ, KS, and MW. All authors read and approved the final manuscript.

Funding

This work was supported by funding from the Medical Research Council (grant: MR/K01353X/1–2). XZ was supported by the Academy of Medical Sciences-Newton Advanced Fellowship (NAFR13\1033). MF and XZ received support from the King's-Peking University Health Science Center Joint Institute for Medical Research. KS was supported by the Biomedical Research Program at Weill Cornell Medicine in Qatar, a programme funded by the Qatar Foundation. TwinsUK is funded by the Wellcome Trust, Medical Research

Council, European Union, Chronic Disease Research Foundation (CDRF), and 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.

Availability of data and materials

Data generated during the study are available as supplementary material. Data on TwinsUK [7] twin participants are available to bona fide researchers under managed access due to governance and ethical constraints. Raw data should be requested via our website (<http://twinsuk.ac.uk/resources-for-researchers/access-our-data/>) and requests are reviewed by the TwinsUK Resource Executive Committee (TREC) regularly. Consent of QMDiab participants does not include public posting of genomics data.

Declarations

Ethics approval and consent to participate

This study was carried out under TwinsUK BioBank ethics, approved by Northwest—Liverpool Central Research Ethics Committee (REC reference 19/NW/0187), IRAS ID 258513, and under the ethics approved by the institutional review boards of Hamad Medical Corporation (HMC) and Weill Cornell Medicine—Qatar (WCM-Q) (research protocol #11131/11). Written informed consent was obtained from all participants. All research was carried out in accordance with the Helsinki Declaration.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 13 January 2024 Accepted: 26 July 2024

Published online: 09 August 2024

References

- Shen X, et al. Multivariate discovery and replication of five novel loci associated with Immunoglobulin G N-glycosylation. *Nat Commun.* 2017;8:447.
- Klarić L, et al. Glycosylation of immunoglobulin G is regulated by a large network of genes pleiotropic with inflammatory diseases. *Sci Adv.* 2020;6:eax0301.
- Gudelj I, Lauc G, Pezer M. Immunoglobulin G glycosylation in aging and diseases. *Cell Immunol.* 2018;333:65–79.
- Ding L, Chen X, Cheng H, Zhang T, Li Z. Advances in IgA glycosylation and its correlation with diseases. *Front Chem.* 2022;10:974854.
- Plomp R, et al. Hinge-Region O-Glycosylation of Human Immunoglobulin G3 (IgG3). *Mol Cell Proteomics.* 2015;14:1373–84.
- Reily C, Stewart TJ, Renfrow MB, Novak J. Glycosylation in health and disease. *Nat Rev Nephrol.* 2019;15:346–66.
- Verdi S, et al. TwinsUK: The UK Adult Twin Registry Update. *Twin Res Hum Genet.* 2019;22:523–9.
- Mook-Kanamori DO, et al. 1,5-anhydroglucitol in saliva is a noninvasive marker of short-term glycemic control. *J Clin Endocrinol Metab.* 2014;99:E479–83.
- Suhre K, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun.* 2017;8:14357.
- Dotz V, et al. O- and N-glycosylation of serum immunoglobulin A is associated with IgA nephropathy and glomerular function. *J Am Soc Nephrol.* 2021;32:2455–65.
- Plomp R, et al. Comparative glycomics of immunoglobulin A and G from saliva and plasma reveals biomarker potential. *Front Immunol.* 2018;9:2436.
- Momčilović A, et al. Simultaneous immunoglobulin A and G glycopeptide profiling for high-throughput applications. *Anal Chem.* 2020;92:4518–26.

13. Ruhaak LR, et al. Targeted biomarker discovery by high throughput glycosylation profiling of human plasma alpha1-antitrypsin and immunoglobulin A. *PLoS ONE*. 2013;8:e73082.
14. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw*. 2015;67(1):1–48. <https://doi.org/10.18637/jss.v067.i01>.
15. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity*. 2005;95:221–7.
16. Scheike TH, Holst KK, Hjelmborg JB. Estimating heritability for cause specific mortality based on twin studies. *Lifetime Data Anal*. 2014;20:210–33.
17. Benedetti E, et al. Network inference from glycoproteomics data reveals new reactions in the IgG glycosylation pathway. *Nat Commun*. 2017;8:1483.
18. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet*. 1998;62:1198–211.
19. Das S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48:1284–7.
20. the Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48:1279–83.
21. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods*. 2012;9:179–81.
22. Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat Methods*. 2014;11:407–9.
23. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 2015;31:3555–7.
24. Welter D, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 2014;42:D1001–6.
25. Kiryluk K, et al. Genome-wide association analyses define pathogenic signaling pathways and prioritize drug targets for IgA nephropathy. *Nat Genet*. 2023;55:1091–105.
26. Ochoa D, et al. The next-generation Open Targets Platform: reimaged, redesigned, rebuilt. *Nucleic Acids Res*. 2023;51:D1353–9.
27. Pickrell JK, et al. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet*. 2016;48:709–17.
28. Pickrell JK. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am J Hum Genet*. 2014;94:559–73.
29. Bladergroen MR, et al. Automation of high-throughput mass spectrometry-based plasma N-Glycome analysis with linkage-specific sialic acid esterification. *J Proteome Res*. 2015;14:4080–6.
30. Paton B, Suarez M, Herrero P, Canela N. Glycosylation biomarkers associated with age-related diseases and current methods for glycan analysis. *IJMS*. 2021;22:5788.
31. Understanding Society Scientific Group, et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet*. 2017;49:416–25.
32. International COPD Genetics Consortium, et al. Genetic overlap of chronic obstructive pulmonary disease and cardiovascular disease-related traits: a large-scale genome-wide cross-trait analysis. *Respir Res*. 2019;20:64.
33. Pucic M, et al. Changes in plasma and IgG N-glycome during childhood and adolescence. *Glycobiology*. 2012;22:975–82.
34. Wang J, et al. Fc-Glycosylation of IgG1 is Modulated by B-cell Stimuli. *Mol Cell Proteomics*. 2011;10:M110.004655.
35. Štambuk J, et al. Global variability of the human IgG glycome. *Aging*. 2020;12:15222–59.
36. Mayboroda OA, Lageveen-Kammeijer GSM, Wuhler M, Dolhain RJEM. An integrated glycosylation signature of rheumatoid arthritis. *Biomolecules*. 2023;13:1106.
37. De Jong SE, et al. IgG1 Fc N-glycan galactosylation as a biomarker for immune activation. *Sci Rep*. 2016;6:28207.
38. Martincic K, Alkan SA, Cheate A, Borghesi L, Milcarek C. Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing. *Nat Immunol*. 2009;10:1102–9.
39. Moldoveanu Z, et al. Patients with IgA nephropathy have increased serum galactose-deficient IgA1 levels. *Kidney Int*. 2007;71:1148–54.
40. Kiryluk K, et al. GWAS for serum galactose-deficient IgA1 implicates critical genes of the O-glycosylation pathway. *PLoS Genet*. 2017;13:e1006609.
41. Li M, et al. Identification of new susceptibility loci for IgA nephropathy in Han Chinese. *Nat Commun*. 2015;6:7270.
42. Fu D, et al. ST6GAL1 polymorphisms influence susceptibility and progression of IgA nephropathy in a Chinese Han population. *Immunobiology*. 2020;225:151973.
43. Putcha N, et al. Lower serum IgA is associated with COPD exacerbation risk in SPIROMICS. *PLoS ONE*. 2018;13:e0194924.
44. Jonsson S, et al. Identification of sequence variants influencing immunoglobulin levels. *Nat Genet*. 2017;49:1182–91.
45. Mitchell JS, et al. Genome-wide association study identifies multiple susceptibility loci for multiple myeloma. *Nat Commun*. 2016;7:12050.
46. Swaminathan B, et al. Variants in ELL2 influencing immunoglobulin levels associate with multiple myeloma. *Nat Commun*. 2015;6:7213.
47. Fleming SC, Smith S, Knowles D, Killen A, Self CH. Increased sialylation of oligosaccharides on IgG paraproteins—a potential new tumour marker in multiple myeloma. *J Clin Pathol*. 1998;51:825–30.
48. Glavey SV, et al. The sialyltransferase ST3GAL6 influences homing and survival in multiple myeloma. *Blood*. 2014;124:1765–76.
49. Nelson AM, Carew NT, Smith SM, Milcarek C. RNA splicing in the transition from B cells to antibody-secreting cells: the influences of ELL2, small nuclear RNA, and endoplasmic reticulum stress. *J Immunol*. 2018;201:3073–83.
50. Ali M, et al. The multiple myeloma risk allele at 5q15 lowers ELL2 expression and increases ribosomal gene expression. *Nat Commun*. 2018;9:1649.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.